

Gewichtungsvariation

Kai Arzheimer

1. Einleitung und Fragestellung

den Anfängen der Umfrageforschung werden routinemäßig Gewichtungsprozeduren angewendet, um die Repräsentativität der Stichproben zu „verbessern“ (Kish 1965), d.h. Ergebnisse zu produzieren, die die tatsächlichen Verhältnisse in der Grundgesamtheit besser widerspiegeln sollen als die Analyse der ungewichteten Daten dies gestatten würde. Ob es durch eine Gewichtung tatsächlich zu einer Verbesserung kommt, läßt sich jedoch kaum nachvollziehen. In der kommerziellen Umfrageforschung werden die verwendeten Gewichtungsprozeduren oft als Geschäftsgeheimnisse betrachtet und basieren zumindest in Einzelfällen auf höchst subjektiven Urteilen der Verantwortlichen. Dies gilt insbesondere für den Bereich der kommerziellen Wahlforschung, in dem die Institute darum wetteifern, Prognosen abzugeben die möglichst gut mit dem endgültigen Ergebnis übereinstimmen. Da die wissenschaftliche Sozialforschung die eigentliche Feldarbeit heute zumeist kommerziellen Instituten überläßt, die nach Abschluß der Feldphase einen fertig aufbereiteten Datensatz einschließlich einer oft nur unzureichend dokumentierten Gewichtungsvariable liefern, wird sie, sofern diese Variablen tatsächlich genutzt werden, nolens volens zum Gefangenen dieser Praktiken.¹ Ein gewisse Skepsis gegenüber Gewichtungen ist deshalb weit verbreitet.

Dieses Mißtrauen speist sich aus zwei weiteren Quellen: Zum einen wird das Thema „Gewichtung“ in den Einführungen und Gesamtdarstellungen zur Umfrageforschung immer noch stiefmütterlich behandelt, auch wenn das Urteil von (Kish 1990: 121), daß sich zumeist nicht einmal ein entsprechendes Stichwort im Index finde, so nicht mehr zu halten ist. Dennoch fehlt es nach wie vor an Texten, die sich gezielt an den potentiellen Nutzer von Gewichtungsverfahren wenden. Die große Mehrheit der Beiträge auf diesem Gebiet ist hochspezialisiert und zielt auf ein primär mathematisch-statistisch ausgebildetes Publikum ab. Der bereits 1994 von Gabler et al. herausgegebene Sammelband zur Anwendung von Gewichtungen in der Umfragepraxis hingegen hat bisher leider keinen Nachfolger gefunden und nicht die von den Herausgebern intendierte Diskussion angestoßen. Im Ergebnis werden nach dem Muster mittelalterlicher Gilden weiterhin Faustregeln und Erfahrungswerte informell von einer Forschergeneration zur nächsten tradiert, aber kaum jemals in einer Form festgehalten, die der intersubjektiven Prüfung zugänglich wäre.

¹ Selbst bei großen europäisch-vergleichend angelegten Studien wie dem Eurobarometer werden die Gewichte von den Umfrageinstituten nach einem nicht näher spezifizierten Verfahren errechnet, das eine Anpassung an die Randverteilung bestimmter sozio-demographischer Merkmale (Alter, Geschlecht etc.) sicherstellen soll.

Erschienen in:
Schoen/Rattinger/
Gabriel (Hrsg.):
Vom Interview
zur Analyse.
Baden-Baden
2009, S. 361-388

Der zweite Grund für die weitverbreitete Skepsis gegenüber Gewichtungen besteht darin, daß häufig nur geringfügige Unterschiede zwischen den gewichteten und den ungewichteten Ergebnissen bestehen (Weisberg 2005: 196). In einigen Fällen können sogar selbst transparente und wissenschaftlich korrekte Gewichtungen dazu führen, daß die gewichteten Daten *stärker* von den Verhältnissen in der Grundgesamtheit abweichen als die ungewichteten Daten, wie die detaillierten Untersuchungen von Rothe zum ALLBUS gezeigt haben (Rothe 1994; zusammenfassend noch einmal Bens 2007).² Soll man nun also überhaupt gewichten?

In seiner klassischen Abhandlung „Weighting: Why, When, and How?“ nennt Leslie Kish (1990: 122) vier Gründe für die Nutzung von Gewichten:

1. Gewichte können unterschiedliche Auswahlwahrscheinlichkeiten, die durch die Methode der Stichprobenziehung verursacht werden, kompensieren.
2. Auswahlwahrscheinlichkeiten können für unterschiedliche Gruppen variiert werden, um die Varianz der Stichprobenparameter zu reduzieren. Auch in diesem Fall werden Gewichte berechnet.
3. Wenn im Rahmen einer Meta- oder Längsschnittanalyse Datensätze kombiniert werden, deren Umfang sich sehr stark unterscheidet, können Gewichte dieses Problem kompensieren.
4. Schließlich können Gewichte genutzt werden, um Verzerrungen zu korrigieren, die sich daraus ergeben, daß Mitglieder verschiedener Gruppen in unterschiedlichem Maße bereit sind, sich an Umfragen zu beteiligen.

Gewichte, die aufgrund der ersten drei Überlegungen berechnet werden, bezeichnet man als „Design-“ oder Transformationsgewichte, weil sie im voraus bekannte Eigenschaften des Designs, die die Repräsentativität der Stichprobe in Frage stellen, kompensieren bzw. die Stichprobe so transformieren, daß sich die gleichen Verteilungen wie bei einer reinen Zufallsstichprobe ergeben. Ein typisches Beispiel für ein solches Designgewicht, das für die umfragebasierte Forschung von besonderer Bedeutung ist, ist die Gewichtung mit der Haushaltsgröße bei zweistufigen Zufallsauswahlen: Hier wird zunächst ein Haushalt zufällig ausgewählt und dann innerhalb des Haushalts wiederum eine zufällig ausgewählte Person befragt.³

Auf dieser zweiten Ebene ist die Auswahlwahrscheinlichkeit für eine gegebene Person von der Haushaltsgröße abhängig. Beispielsweise wird eine Person in einem Einpersonenhaushalt mit Sicherheit (100%) ausgewählt, während die Auswahlwahrscheinlichkeit in einem Fünf-Personen-Haushalt nur 1/5 bzw. 20% beträgt. In der Konsequenz sind in der Stichprobe Personen aus Ein-Personenhaushalten (die auch bezüglich einer Reihe anderer Merkmale vom Mittel der Bevölkerung abweichen) überrepräsentiert.

² Eine naheliegende Erklärung für dieses Phänomen ist, daß einigen wenigen Fälle, die eine Gruppe mit sehr geringer Befragungsbereitschaft repräsentieren sollen, sehr hohe Gewichtungsfaktoren zugewiesen werden. Wenn sich diese Fälle stark von jenen Gruppenmitgliedern, die sich nicht an der Umfrage beteiligt haben, unterscheiden, muß es zu substantiellen Verzerrungen kommen.

³ Die Ausführungen gelten sinngemäß auch für drei- und mehrstufige Auswahlen. Kish's Fälle (2) und (3) können hier unberücksichtigt bleiben.

Ein Rechenbeispiel verdeutlicht dies: In einer fiktiven Gesellschaft seien 50% aller Haushalte Ein-, die übrigen 50% Fünf-Personen-Haushalte. Zieht man nun wie oben beschrieben eine zweistufige Stichprobe, wird diese 50% Alleinstehende enthalten. Tatsächlich lebt aber in der fiktiven Gesellschaft nur jede sechste Person alleine, was einem Bevölkerungsanteil von knapp 17 Prozent entspricht. Anders gewendet ist die Stichprobe auf der Meso-Ebene der Haushalte repräsentativ, nicht aber auf der Mikro-Ebene der Individuen, auf die sich die meisten sozialwissenschaftlichen Theorien beziehen. Dieses Problem läßt sich aber ex post beheben, indem man die Fälle mit dem Kehrwert der Auswahlwahrscheinlichkeit der betreffenden Person ($1/0.5=2$ bzw. $1/0.1=10$) gewichtet.⁴

Die Berechnung von Designgewichten ergibt sich logisch aus dem Plan der Stichprobenziehung und ist deshalb unproblematisch und im Grunde unumstritten. In der Praxis zeigt sich allerdings häufig, daß die Verwendung von Designgewichten kaum einen Einfluß auf die Ergebnisse der Analyse hat. Ein möglicher Grund dafür besteht darin, daß diese Gewichte auf der theoretischen Auswahlwahrscheinlichkeit basieren, die für Menschen in Ein-Personen-Haushalten wie gezeigt erheblich größer ist als für Bürger, die in Mehr-Personen-Haushalten leben. Durch die Tatsache, daß Alleinstehende seltener zu Hause anzutreffen sind und größere Hemmungen haben, ihnen unbekanntem Interviewern die Türe zu öffnen, wird diese höhere Auswahlwahrscheinlichkeit aber zum Teil wieder kompensiert.

Der eigentliche Gegenstand der oben beschriebenen Skepsis gegenüber Gewichtungsverfahren ist jedoch nicht die mehr oder minder sinnvolle Designgewichtung, sondern vielmehr solche Gewichtungsvariablen, die gemeinsame Verteilung einiger weniger Variablen in der Stichprobe an deren bekannte gemeinsame Verteilung in der Grundgesamtheit anpassen sollen („Redressment“), um so die Verzerrungen auszugleichen, die sich aus selektiven Antwortausfällen ergeben. Bei dieser am häufigsten verwendeten Gewichtungsprozedur handelt es sich also (auch wenn dies nicht allen Anwendern bewußt sein dürfte) um eine Technik zur Behandlung von Missing-Data-Problemen, die allerdings im Gegensatz zu vielen anderen Verfahren nicht bei den partiellen (item-nonresponse), sondern vielmehr bei den vollständigen Ausfällen (unit-nonresponse) ansetzt. Damit eine Redressement-Gewichtung den gewünschten Erfolg haben kann, müssen allerdings zwei Voraussetzungen erfüllt sein: Die zur Gewichtung verwendeten Variablen müssen erstens in einem engen Zusammenhang zur Teilnahmebereitschaft stehen und zweitens dürfen sich *innerhalb* der durch die Gewichtungsvariablen definierten Gruppen Teilnehmer und Interviewverweigerer nicht systematisch unterscheiden.⁵ Weitere Probleme ergeben sich daraus, daß möglichst viele Variablen, die die Teilnahmewahrscheinlichkeit beeinflussen können, berücksichtigt werden sollten, um die zweite Annahme plausibler zu machen. Schon bei einer relativ geringen Zahl von Gewichtungsvariablen werden jedoch in einer normal dimensionierten Stichprobe viele Zellen der Gewichtungsmatrix nur schwach oder gar nicht besetzt

⁴ In der Praxis müssen diese Gewichte noch so normalisiert werden, daß sich über alle Gewichte ein Mittelwert von 1 ergibt. Andernfalls würde sich die für Signifikanztests relevante Zahl der Fälle verändern. Im oben beschriebenen Fall wären die korrekten Gewichte 0.333 und 1.667.

⁵ Darüber hinaus empfiehlt es sich, nur solche Merkmale zur Gewichtung zu verwenden, deren Verteilung in der Grundgesamtheit nicht nur bekannt, sondern auch langfristig stabil ist, da es sonst durch eine Gewichtung zu massiven Verzerrungen kommen kann (Weisberg 2005: 195).

sein, woraus extreme Gewichtungsfaktoren resultieren, die zu sehr ernsthaften Verzerrungen in der *gewichteten* Stichprobe führen können (vgl. FN).

Zu den häufig geäußerten Faustregeln für die Gewichtung von Umfragedaten gehören die Aussagen, daß a) Gewichtungen dann (und nur dann) sinnvoll sind, wenn die Gewichtungsvariablen in einem engen Zusammenhang mit den für die Studie relevanten Variablen stehen, daß b) Antwortausfälle in erster Linie die Schätzungen für Mittelwerte und Randverteilungen beeinflussen, während die (aus wissenschaftlicher Sicht in der Regel sehr viel interessanteren) für Schätzungen für Zusammenhänge und Regressionskoeffizienten relativ robust seien und daß c) Gewichtsprozeduren kaum einen Einfluß auf die Schätzungen für Zusammenhänge/Koeffizienten haben.

Daß die erste dieser Aussagen zutreffen muß, ist im Grunde eine Trivialität: Wenn zwischen einer Gewichtungsvariablen a und den untersuchten Variablen x und y kein substantieller Zusammenhang besteht, wird eine Gewichtung nach a die Verteilungen (Varianzen und Mittelwerte) von x und y in der Stichprobe nicht (systematisch) beeinflussen. Dementsprechend werden sich auch gewichtete und ungewichtete Ergebnisse nicht substantiell unterscheiden.

Schwieriger ist es, die zweite und die dritte Aussage zu beurteilen: In welchem Umfang Zusammenhänge (d.h. in erster Linie die Koeffizienten einer linearen Regression) durch unit-nonresponse beeinflusst werden, hängt davon ab, welche Beziehungen zwischen x , y und der bzw. den Variablen, die für einen Antwortausfall verantwortlich sind, bestehen.

Ein Computer-Experiment verdeutlicht dies. In einer Grundgesamtheit soll der folgende Zusammenhang bestehen: $y=3+0.7*x+\varepsilon$. Die Varianz von ε beträgt 20.25. x ist mit einer Kurtosis von 4 und einer skewness von 0.64 leicht breitgipflig und etwas rechtsschief, während y normalverteilt ist. Der Mittelwert und die Varianz von x liegen bei 0.3 und 12.8; die entsprechenden Werte für y liegen bei 3.2 und 26.5. Die Wahrscheinlichkeit eines Antwortausfalls soll von x beeinflusst werden. Dabei wurden eher optimistische Annahmen getroffen: Die Wahrscheinlichkeit eines Ausfalls ist bei niedrigen x -Werten sehr hoch, sinkt dann aber rasch ab (vgl. Abbildung 1). Am unteren Quartil von x (-2) liegt die Ausfallwahrscheinlichkeit bei lediglich rund 40 Prozent, am Median (0) bei 22 Prozent und am oberen Quartil (2.3) bei weniger als zehn Prozent (vgl. Abbildung 1). Insgesamt beträgt die Wahrscheinlichkeit eines Ausfalls 28 Prozent, was angesichts der in der Praxis beobachteten Ausfälle von 40 Prozent und mehr sehr niedrig angesetzt ist. Zur Gewichtung steht eine Variable a mit ganzzahligen Ausprägungen von 1 bis 5 zur Verfügung, die moderat mit y und x korreliert ist (0.11/0.24). Um das Beispiel etwas anschaulicher zu gestalten, kann man sich vorstellen, daß es sich bei x um das generelle politische Interesse und bei y um einen Indikator für tatsächliche politische Partizipation handeln könnte. Bei a könnte es sich um eine Variable handeln, die die formale Bildung erfaßt.

Aus dieser fiktiven Grundgesamtheit wurden nun in einer Computersimulation 1000 Stichproben vom Umfang $n=2000$ gezogen. Die Abbildungen 2-7 zeigen die Ergebnisse. Dabei stehen die durchgezogenen Linien jeweils für unverzerrten Zufallstichproben aus der Grundgesamtheit. Die gestrichelten Linien zeigen die Verteilung der Mittelwerte aus Stichproben, die von dem oben skizzierten

Ausfallmechanismus betroffen sind (und dementsprechend auch einen geringeren Umfang haben), die gepunktete Linie schließlich zeigt die Verteilung, nachdem eine Gewichtung durchgeführt wurde.

Da die Ausfallwahrscheinlichkeit vom Wert von x abhängt, sind die Schätzungen für die Mittelwerte von x und y erwartungsgemäß verzerrt: Für x liegt der Erwartungswert des Mittelwertes etwa einen, für y rund 0.7 Skalenpunkte zu hoch. Dabei spielte es substantiell keine Rolle, ob die verzerrte Stichprobe gewichtet wurde. Ob eine solche Verzerrung die inhaltlichen Schlüsse, die aus den Ergebnissen der Analyse gezogen werden beeinflussen würde, läßt sich im Kontext eines solchen Experiments nicht beantworten, da dies naturgemäß von der Fragestellung und der inhaltlichen Bedeutung der betroffenen Variablen abhängt. Gemessen an der Standardabweichung von x und y (3.6/5.1) sind die Abweichungen vom wahren Mittelwert jedoch relativ klein.

Allerdings kommt es nicht nur bei der Schätzung der Mittelwerte, sondern auch bei der Schätzung der Varianzen zu Verzerrungen. Dies gilt vor allem für die Varianz von x , die um rund 12 Prozent unterschätzt wird, da Fälle mit sehr niedrigen x -Werten in den Stichproben deutlich unterrepräsentiert sind. Problematisch ist dies vor allem deshalb, weil bei der Berechnung von Standardfehlern die unbekannte Varianz von x in der Grundgesamtheit durch die Varianz von x in der Stichprobe geschätzt wird. Im bivariaten Fall ist der Standardfehler umgekehrt proportional zur Varianz von x , d.h. zu konservativ. Zudem hat die reduzierte Varianz von x auch einen negativen Einfluß auf die Berechnung von R^2 , das trotz der bekannten Probleme (Achen 1977; King 1986) von vielen Forschern immer noch als Maß für die Qualität eines Regressionsmodells betrachtet wird. Konkret wird in den von Ausfällen betroffenen Stichproben statt des korrekten Wertes von 0.23 im Mittel nur ein Wert von 0.21 erreicht (nicht grafisch ausgewiesen). Die Anwendung des Gewichtungsverfahrens hat auf dieses Ergebnis wiederum faktisch keinen Einfluß.

Auf die *Parameter* des Regressionsmodells scheinen die simulierten Ausfälle allerdings keinerlei Einfluß zu haben. Abbildung 6 zeigt die Verteilungen für die Schätzungen der Konstante. Die Mittelwerte aller drei Verteilungen sind bis auf drei Nachkommastellen identisch und entsprechen mit einem Wert 3.005 fast exakt dem wahren Wert in der Grundgesamtheit. Dasselbe gilt auch für die Verteilungen der Schätzungen für die Steigung der Regressionsgeraden (vgl. Abbildung 7).⁶ Diese Ergebnisse scheinen die Faustregeln b) und c) zu bestätigen: Trotz erheblicher systematischer Ausfälle sind die Schätzungen für den Zusammenhang zwischen x und y unverzerrt, unabhängig davon ob gewichtet wird oder nicht.

Tatsächlich geht dieser erfreuliche Befund jedoch auf den Ausfallmechanismus zurück, der ja so spezifiziert wurde, daß die Wahrscheinlichkeit einer unit-nonresponse ausschließlich vom Wert von x , nicht aber vom Wert von y beeinflusst wird. In der Realität kann es aber durchaus der Fall sein, daß höhere Werte von x zwar die Ausfallwahrscheinlichkeit reduzieren, *ceteris paribus* aber mit höheren Werten von y die Wahrscheinlichkeit einer unit-nonresponse steigen. Mit Blick auf die oben gewählte inhaltliche Interpretation des Beispiels erscheint dies übrigens durchaus plausibel: Hält man das politische Interesse (das einen positiven Effekt auf

⁶ Die etwas größere Streuung der Koeffizienten aus den von Ausfällen betroffenen Stichproben erklärt sich aus deren geringerem Umfang.

die Bereitschaft zur Teilnahme an Umfragen hat) konstant, so dürfte mit zunehmender politischer Aktivität die Wahrscheinlichkeit der Teilnahme *sinken*, da die prospektiven Befragten schlicht weniger Zeit zur Verfügung haben. Abbildung 8 illustriert die Verhältnisse.

Die Auswirkungen dieser scheinbar harmlosen Annahme sind bemerkenswert. Die Abbildungen 9 und 10 zeigen zunächst, daß die Schätzungen für Mittelwert und Varianz von x weiterhin verzerrt sind, allerdings weniger stark als unter dem ersten Ausfallmechanismus. Wie im ersten Szenario hat die Gewichtung keinen substantiellen Einfluß auf das Ausmaß der Verzerrung. Der Mittelwert von y wird nun deutlich *unterschätzt* (Abbildung 11), und auch im Fall der Varianz von y ist die Verzerrung nun stärker als zuvor (Abbildung 12). Wiederum ist es völlig unerheblich, ob die Daten gewichtet werden oder nicht. Geradezu dramatisch sind allerdings die Auswirkungen auf die Parameter des Regressionsmodells: Unabhängig davon, ob gewichtet wird oder nicht liegen der Median und das arithmetische Mittel für die Schätzung der Konstanten bei 1.29, d.h. weit unter dem wahren Wert von 3 (Abbildung 13). Ein ganz ähnliches Bild ergibt sich für den Steigungsparameter: Hier wird statt des wahren Wertes von 0.7 unabhängig vom Gewichtungsverfahren ein Wert von 0.88 geschätzt. Da die Varianz von x beträchtlich ist (in der fiktiven Grundgesamtheit werden Werte zwischen -11 und +18 beobachtet), wird der Effekt von x auf y deutlich überschätzt. Noch eindrucksvoller ist der Effekt der systematischen Ausfälle auf den Wert von R^2 . Statt des wahren Wertes von 0.23 wird in den Stichproben im Mittel ein Wert von über 0.40 erreicht. Der kleinste in den Stichproben errechnete Wert liegt mit 0.33 immer noch weit über dem Wert in der Grundgesamtheit.

Aus den Ergebnissen der Simulation lassen sich zwei Schlußfolgerungen ableiten: Zum einen läßt sich die generelle Aussage, daß Zusammenhänge bzw. Regressionskoeffizienten gegenüber systematischen Ausfällen unempfindlich sind, in dieser Form leider nicht halten. Vielmehr gilt dies (im bivariaten Fall) nur dann, wenn die Ausfallwahrscheinlichkeit ausschließlich vom Wert der unabhängigen Variablen abhängt (Allison 2002). Zum anderen bestätigen die Ergebnisse, einmal mehr daß eine Gewichtung häufig keinen substantiellen Einfluß auf die Ergebnisse hat, sofern nicht ein sehr enger Zusammenhang zwischen den Gewichtungsvariablen und den analysierten Variablen besteht. Selbst in diesem Fall sind Gewichtungen allerdings häufig entbehrlich, da die Variablen, die üblicherweise zur Gewichtung verwendet werden – Alter, Geschlecht, formale Bildung, Region – meist ohnehin aus inhaltlichen Gründen oder als sogenannte Kontrollvariablen mit in das Regressionsmodell aufgenommen werden, so daß eine Anpassung der Randverteilungen zumeist nur einen geringen Einfluß auf die Ergebnisse haben wird.

Dies gilt in der Tendenz auch für das vorliegende Projekt. Obwohl die Zahl der Gewichtungsvariablen aufgrund der Komplexität des Studiendesigns sehr groß ist – der kombinierte Querschnittsdatensatz 1994-2002 enthält fünf, der kombinierte Paneldatensatz 1994-2002 nicht weniger als siebzehn Gewichtungsvariablen – lassen sich diese im wesentlichen auf zwei Typen reduzieren: Eine Reihe von Ost-West-Gewichten trägt der Tatsache Rechnung, daß die neuen Bundesländer aus inhaltlichen Gründen im Sample überrepräsentiert sind, während eine Serie von

Repräsentativgewichten mit dem Ziel berechnet wurden, die Verteilung einiger sozio-demographischer Merkmale an die bekannten Verteilungen in der Bevölkerung anzupassen. Für die Jahre 1994 und 1998 wurden diese Gewichte von Emnid errechnet und beziehen sich auf die Anzahl Personen im Haushalt, das Geschlecht, das Alter, das Bundesland und die politische Ortsgrößenklasse. Für 2002 wurden die Gewichte auf Grundlage derselben Variablen von INRA errechnet. Zumindest das Alter und das Geschlecht werden von vielen Forschern routinemäßig in ihre Regressionsmodelle aufgenommen. Im Kontext der vorliegenden Studie gilt das auch für die Region.

Beide Gewichtungstypen wurden außerdem multiplikativ zu einer gemeinsamen Variablen verknüpft, deren Verwendung gesamtdeutsch-repräsentative Auswertungen erlauben soll. Hinzu kommen noch einige besondere Gewichte, die beispielsweise allen drei Befragungswellen eine identische Fallzahl zuweisen und hier außer Betracht bleiben können. Die sehr große Zahl von Gewichtungsvariablen im Paneldatensatz ergibt sich daraus, daß einerseits (wie auch im Querschnitt) Gewichte getrennt für die Vor- und Nachwahlbefragungen und dann noch einmal gemeinsam für beide Befragungsteile berechnet wurden, andererseits für verschiedene Kombinationen von Panelwellen – beispielsweise 1994/2002 oder 1994/1998/2002 separate Gewichtungsvariablen bereitgestellt werden.⁷

Um einen Eindruck von der substantiellen Bedeutung der Gewichte zu bekommen, ist es weder notwendig noch sinnvoll, auf alle verfügbaren Variablen zurückzugreifen. Im folgenden werden vielmehr stets drei Szenarien miteinander verglichen: die Analyse der ungewichteten Daten, eine reine Ost-West-Gewichtung sowie die Kombination von Ost-West- und (regionsspezifischer) Repräsentativgewichtung. Auf eine Aufspaltung der Befragungswellen in Vor- und Nachwahluntersuchungen wird dabei generell verzichtet. Im Falle der Paneldaten werden nur jene Befragten untersucht, die sich an allen drei Wellen beteiligt haben, da hier aufgrund der Panelmortalität die stärksten Verzerrungen zu erwarten sind.

Anders als in vielen primär methodologischen Beiträgen sollen hier nicht vollständige Kovarianzmatrizen miteinander verglichen werden, da sich in solchen Matrizen eine Vielzahl von (häufig trivialen) Abweichungen nachweisen läßt, deren inhaltliche Bedeutung aber völlig unklar bleibt. Statt dessen werden lineare Regressionsmodelle geschätzt, die in der Forschungspraxis noch immer die mit Abstand wichtigste Rolle spielen. Als „Material“ für die drei Gewichtungsvarianten dienen dabei zwei einfache Modelle, die sich an bekannte Zusammenhänge aus der politischen Soziologie anlehnen. Das erste Modell definiert auf der Basis zweier Einstellungsfragen (Bereitschaft zur Mitarbeit in einer Bürgerinitiative und zur Teilnahme an einer genehmigten Demonstration) einen einfachen Summenindex „unkonventioneller“ (Barnes et al. 1979) politischer Partizipation.⁸ Dieser Index wurde auf eine Reihe von potentiell erklärungskräftigen Variablen wie den Inglehart-Index, das Vorhandensein einer Parteiidentifikation, die wahrgenommene

⁷ Tatsächlich ist die Zahl der Gewichtungsmöglichkeiten größer als die der Gewichtungsvariablen, da die für verschiedene Prozeduren benötigten Informationen nach Möglichkeit in einer Variablen zusammengeführt wurden.

⁸ Sofern eines der beiden Items von einem Antwortausfall betroffen war, wurde der Wert des anderen Items zur Berechnung des Index verdoppelt, um möglichst viele Fälle nutzen zu können.

Performanz des politischen Systems sowie die eigene Position im Links-Rechts-Spektrum⁹ regrediert. Alle diese Variablen sollten einen positiven Einfluß auf die politische Partizipation haben. Zudem wurde eine Reihe von Kontrollvariablen (Alter, Geschlecht, formale Bildung, Region) berücksichtigt.

Das zweite Modell ist etwas komplexer und lehnt sich stärker an konkrete Befunde aus dem Bereich der Forschung zum Institutionenvertrauen an (Rohrschneider/ Schmitt-Beck 2002). Die abhängige Variable ist hier der Mittelwert¹⁰ des mittels der gängigen Ratingskalen erhobenen Vertrauens in Bundestag, Bundesregierung und Bundesverfassungsgericht. Als Kontrollvariablen dienen wiederum Alter, Bildung und Geschlecht und Region. Da sich in der Vergangenheit gezeigt hat, daß das Vertrauen in die politischen Kerninstitutionen von der (wahrgenommenen) Leistung des politischen Systems beeinflusst wird, wurden neben der bereits bekannten Unzufriedenheit mit dem Funktionieren der Demokratie in Deutschland auch zwei Indikatoren zur Wahrnehmung der allgemeinen bzw. der persönlichen Wirtschaftslage in das Modell aufgenommen. Darüber hinaus steht zu erwarten, daß Anhänger der Regierung ein höheres Vertrauen in die politischen Institutionen haben. Deshalb wurde ein entsprechender Dummy gebildet, der für Personen, die angaben, für eine der (bisherigen) Regierungsparteien gestimmt zu haben bzw. stimmen zu wollen den Wert 1 annimmt. Abschließend wurden noch zwei Indikatoren mit in das Modell aufgenommen, die im Ansatz eine *grundsätzliche* Opposition zum bestehenden politischen System, d.h. eine Präferenz für ein alternatives Regime erfassen: die Aussage, daß die DDR „mehr gute als schlechte Seiten“ bzw. daß der Nationalsozialismus „auch seine guten Seiten“ gehabt habe.

2. Empirische Ergebnisse: Querschnitt

Tabelle 1 zeigt die Ergebnisse für das Partizipationsmodell. Aus Gründen der Übersichtlichkeit wurden die Daten für alle drei Befragungswellen gemeinsam analysiert. Eine jahresweise Analyse der Daten (nicht tabellarisch ausgewiesen) führt aber zu substantiell identischen Ergebnissen. Die linke Spalte enthält die Schätzungen, die auf den ungewichteten Daten basieren. Bei den Einträgen handelt es sich um unstandardisierte Regressionskoeffizienten, die sich auf den oben vorgestellten Partizipationsindex mit einer Spannweite von 2 bis 10 Skalenpunkten beziehen.

Alle Zusammenhänge entsprechen den Erwartungen: Verglichen mit Materialisten sind Mischtypen und vor allem Postmaterialisten weitaus eher bereit, sich „unkonventioneller“ politischer Mittel zu bedienen. Die Differenz zwischen Postmaterialisten und Materialisten (der Referenzgruppe) beträgt mehr als 1.6 Skalenpunkte. Eine Links-Rechts-Selbsteinstufung, die stark vom analytischen Mittelpunkt der Skala abweicht, hat ebenfalls einen nennenswerten Effekt auf die

⁹ Verwendet wird der Betrag der Abweichung vom Mittelpunkt der Skala; d.h. der Wert ist stets positiv und um so höher, je radikaler (links oder rechts) die Position ist, die ein Befragter einnimmt.

¹⁰ Um möglichst viele Interviews analysieren zu können, wurden in diesen Mittelwert auch Fälle einbezogen, in denen nur zwei oder sogar nur eine Institution beurteilt wurde.

Partizipationswahrscheinlichkeit, der allerdings schwächer ist als der Einfluß der Wertorientierungen: mit jedem Skalenpunkt auf der Links-Rechts-Achse ist eine Zunahme von 0.16 Punkten auf der Partizipationskala verbunden. Der maximale Effekt der ideologischen Selbstverortung beträgt dementsprechend 0.8 Skalenpunkte. Auch das Vorhandensein einer (beliebigen) Parteiidentifikation hat einen moderaten positiven Einfluß auf die Partizipationswahrscheinlichkeit; sei es, weil Parteien in den Netzwerken der sozialen Bewegungen weiterhin eine wichtige Rolle spielen oder weil Parteiidentifizierer generell ein höheres Maß an politischer Aktivität zeigen als Nicht-Identifizierer. Die Unzufriedenheit mit dem Funktionieren des politischen Systems schließlich hat nur einen geringen positiven Einfluß auf das Niveau der unkonventionellen politischen Partizipation: die Differenz zwischen sehr zufriedenen und sehr unzufriedenen Bürgern beträgt ebenfalls 0.4 Skalenpunkte.

Auch die Effekte der Kontrollvariablen entsprechen weitgehend den Erwartungen: Jüngere Menschen (bis 45 Jahre) haben ein größeres Interesse an unkonventionellen Partizipationsformen als ihre zwischen 45 und 65 Jahre alten Mitbürger. Diese wiederum sind deutlich aktiver als die Referenzgruppe der über 65 Jahre alten Befragten. Unabhängig von diesen Alterseffekten haben mittlere und vor allem höhere Bildungsabschlüsse einen sehr deutlichen positiven Effekt, während das Geschlecht einen signifikanten, aber deutlich schwächeren Einfluß hat. Werden alle diese Variablen statistisch kontrolliert, läßt sich schließlich kein signifikanter Ost-West-Unterschied mehr feststellen.

In einem nächsten Analyseschritt wurde nun die reine Ost-West-Gewichtung auf die Daten angewendet. Während die Ostdeutschen über den gesamten Untersuchungszeitraum hinweg etwa ein Fünftel der gesamtdeutschen Bevölkerung ausmachten, lag auf Grund des Stichprobendesigns ihr Anteil an den Befragten der ersten Welle bei knapp 50 Prozent. In den beiden folgenden Befragungswellen wurde der Anteil der Ostdeutschen an den Befragten auf etwa ein Drittel reduziert. Aus diesen Zahlenverhältnissen ergeben sich relativ moderate¹¹ Gewichtungsfaktoren im Bereich von 0.5 für Ostdeutschland und 1.3 für Westdeutschland.¹²

Die mittlere Tabellenspalte enthält die Ergebnisse. Bereits auf den ersten Blick ist zu erkennen, daß diese Form der Gewichtung für die substantielle Bewertung des Modells völlig unerheblich ist. Die Koeffizienten unterscheiden sich nur geringfügig, und alle Differenzen bewegen sich durchgängig im Bereich von weniger als zwei Standardfehlern. Die größte Abweichung von 0.16 Skalenpunkten – auch dies ist mit Blick auf die Skalierung der abhängigen Variablen eine völlig irrelevante Differenz – zeigt sich bei der Konstanten.

Dramatischere Unterschiede sind von der kombinierten Ost-West-/Repräsentativgewichtung zu erwarten. Die entsprechende Variable „vgges“ enthält 3189 verschiedene Werte zwischen 0.114 und 7.79. Somit wird den Fällen mit dem höchsten Faktor – zwei jungen Frauen aus der letzten Befragungswelle, die

¹¹ Trotzdem wird den Befragten mit dem höchsten Gewichtungsfaktor (Westdeutsche 1994) fast viermal soviel Gewicht zugemessen wie jenen mit dem niedrigsten Faktor (Ostdeutsche 1994).

¹² Die entsprechende Variable „vgwow2“ enthält insgesamt acht verschiedene Werte, wobei die vier Faktoren für 1998 und 2002 nahezu identisch sind. 1994 wurden für die Vor- und die Nachwahlbefragung jeweils getrennte Werte berechnet, die aber wiederum fast identisch sind.

in einem Zwei- bzw. einem Vier-Personen-Haushalt in Saarbrücken bzw. einer Mecklenburg-Vorpommerschen Kleinstadt leben – rund 67 mal soviel Gewicht verliehen wie den drei Befragten mit dem niedrigsten Gewichtungsfaktor (drei alleinlebenden Männer aus Saarländischen Mittelstädten, die in der ersten Befragungswelle interviewt wurden). Sicherlich handelt es sich hierbei um ein besonders extremes Beispiel. Angesichts der Verteilung von v_{gges} (Abbildung 15) stellt sich jedoch die Frage, ob es tatsächlich vertretbar ist, einem bestimmten Fall mit einer ungewöhnlichen Kombination von sozio-demographischen Merkmalen ein sieben bis zehnmal höheres Gewicht zu geben als einem anderen Befragten.

In der Praxis haben aber selbst diese extremen Gewichtungsfaktoren keinen substantiellen Einfluß auf das Ergebnis der Regressionsanalyse. Vergleicht man die erste und die letzte Spalte von Tabelle 1, so ist wiederum zu erkennen, daß die Koeffizienten für alle praktischen Zwecke so gut wie identisch sind. Die größte Differenz von 0.15 Skaleneinheiten zeigt sich nun beim Koeffizienten für die jüngste Altersgruppe. In der großen Mehrzahl der übrigen Fälle sind die Ergebnisse für die ungewichteten und die Ost-West-/repräsentativgewichteten Daten weniger stark voneinander als die Ergebnisse für die ungewichteten bzw. ausschließlich Ost-West-gewichteten Daten, wobei die Diskrepanzen zwischen beiden Gewichtungsvarianten substantiell zu vernachlässigen sind.

Eine naheliegende Erklärung für diesen Befund könnte darin bestehen, daß ein Teil der für die Gewichtung verwendeten Merkmale bereits als Kontrollvariablen im Modell enthalten ist. Schließt man diese Größen jedoch aus dem Modell aus, so ergeben sich wiederum keine substantiellen Abweichungen zwischen den drei Gewichtungsvarianten (vgl. Tabelle 2).

Tabelle 3 zeigt die Ergebnisse für das zweite Modell, in dem das Vertrauen in die politischen Institutionen auf eine Reihe von Einstellungsvariablen regrediert wird. Auch hier enthält die linke Spalte die Ergebnisse für die ungewichteten Daten, in der mittleren Spalte sind die Koeffizienten eingetragen, die auf Grundlage der regionalgewichteten Daten errechnet wurden, während die rechte Spalte die Ergebnisse der kombinierten Gewichtung zeigt. Wiederum entsprechen die Ergebnisse im wesentlichen den theoretischen Erwartungen. Anhänger der Regierungsparteien weisen im Schnitt ein um 0.4 Punkte höheres Vertrauen in die politischen Institutionen auf. Berücksichtigt man die Skalierung der abhängigen Variablen, die Werte zwischen 1 und 5 annehmen kann, ist dies ein relevanter Zusammenhang. Eine negative Wahrnehmung der individuellen und der allgemeinen Wirtschaftslage haben beide den erwarteten negativen Effekt auf das Vertrauen in die Institutionen. Die Koeffizienten sind zwar relativ klein, jedoch ist angesichts ihrer Spannweite (1 bis 5) der maximale Effekt beider Variablen nicht zu vernachlässigen.

Deutlich stärker ist der maximale Effekt der von einer negativen Wahrnehmung der Performanz des politischen Systems ausgeht: Die Differenz zwischen sehr zufriedenen und sehr unzufriedenen Bürgern beträgt rund einen Skaleneinheit. Die Effekte der Wertorientierungen, der ideologischen Selbsteinstufung sowie der Unterstützung für ein alternatives Regime (Sozialismus in der DDR bzw. Nationalsozialismus) entsprechen hingegen von ihrer Richtung her zwar den

theoretischen Erwartungen, sind aber sehr schwach und inhaltlich fast bedeutungslos.

Auch die Kontrollvariablen haben (fast) keinen Einfluß. Statistisch signifikante (aber inhaltlich bedeutungslose) Effekte zeigen sich nur für das Geschlecht, die beiden jüngeren Altersgruppen sowie die Region. Deren (schwacher) Effekt ist erstaunlicherweise positiv, d.h. Ostdeutsche zeigen *ceteris paribus* ein geringgradig höheres Institutionenvertrauen als ihre westdeutschen Mitbürger. Dieser kontraintuitive Befund, der im Widerspruch zu den bisherigen Forschungsergebnissen in diesem Feld steht, erklärt sich jedoch aus der großen Zahl von Einstellungsvariablen, die das Modell enthält: Im Osten Deutschlands leben weniger Anhänger der Regierung (-5 Prozentpunkte), die Unzufriedenheit mit dem Funktionieren der Demokratie in Deutschland ist größer (0.5 Skalenpunkte), die individuelle und vor allem die kollektive Wirtschaftslage werden deutlich negativer wahrgenommen (0.1/0.3 Skalenpunkte) und die Zahl derjenigen, die die untergegangene DDR positiv bewerten, ist dramatisch höher (1.2 Skalenpunkte). Werden alle diese Faktoren statistisch kontrolliert, ist der verbleibende Regionaleffekt schwach positiv.¹³

Auf Grundlage der ungewichteten Daten würde man schließen, daß das Institutionenvertrauen im vereinten Deutschland in erster Linie von der wahrgenommenen Leistungsfähigkeit des politischen Systems sowie der wahrgenommenen Wirtschaftslage abhängt. Zudem gewähren Anhänger der Regierungsparteien den Institutionen einen zusätzlichen Vertrauensbonus. Ideologische Einstellungen scheinen hingegen nur einen relativ geringen Einfluß auf das Vertrauen in die politischen Institutionen zu haben.

Ergibt sich ein substantiell anderes Bild, wenn die Daten gewichtet werden? Ein Vergleich der linken und der mittleren Spalte von Tabelle 3 zeigt, daß dies nicht der Fall ist. Die Abweichungen zwischen ungewichteten und ost-west-gewichteten Daten sind vielmehr noch kleiner als im Falle des ersten Modells. Für eine große Zahl von Variablen stimmen die Koeffizienten bis auf zwei Nachkommastellen überein. In keinem Fall beträgt die Differenz zwischen gewichteten und ungewichteten Koeffizienten mehr als 0.1 Skalenpunkte.

Auf Basis der umfassenderen Ost-West-/Repräsentativgewichtung ergeben sich wiederum Ergebnisse, die zumeist noch näher an den Koeffizienten liegen, die für die ungewichteten Daten berechnet wurden. Substantiell sind alle Differenzen zwischen ungewichteten, ost-west-gewichteten und ost-west-/repräsentativgewichteten Daten vollkommen irrelevant.

3. Empirische Ergebnisse: Panel

Wie in anderen Kapiteln dieses Bandes ausführlich dargelegt, ist das Design der vorliegenden Studie außerordentlich komplex. Innerhalb der Panel-Komponente wurde ein Teil der Befragten bis zu drei Mal (1994, 1998 und 2002) befragt, wobei das Panel 1998 aufgefüllt wurde, um die unvermeidliche Panelmortalität zu kompensieren. Für die Analysen dieses Abschnitts wird ausschließlich auf jene mit n=1423 recht kleine Gruppe von Befragten zurückgegriffen, die sich an allen drei

¹³ Dennoch bleibt festzuhalten, daß die hier beobachteten Ost-West-Differenzen (0.2 Skalenpunkte über alle drei Wellen hinweg) verglichen mit dem ALLBUS und anderen Studien relativ klein sind.

Befragungswellen beteiligt haben, da es sich hier ganz offensichtlich um einen Fall von Selbstselektion handelt, der mit entsprechenden starken Abweichungen gegenüber der Grundgesamtheit einhergehen sollte.

Tatsächlich zeigt sich, daß die Gruppe der Panelteilnehmer eine sehr ungewöhnliche Zusammensetzung hat: Der Anteil der Männer ist mit 55 Prozent deutlich höher als in der Grundgesamtheit, der Anteil der formal Hochgebildeten liegt bei fast 35 und der Anteil der Ostdeutschen bei über 58 Prozent. Auch der Anteil der Parteiidentifizierer liegt mit 79 Prozent weit über dem Wert in der Grundgesamtheit (Arzheimer/ Schoen 2005). Dementsprechend steht zu erwarten, daß die Verwendung von GewichtungsvARIABLEN hier einen deutlich größeren Einfluß auf die inhaltlichen Ergebnisse haben sollte als im Falle der in Abschnitt 2 analysierten Querschnittsdaten. Die Verteilung der GewichtungsvARIABLEN selbst ist erwartungsgemäß noch extremer als im Querschnitt (vgl. Abbildung 16). Dem Fall mit dem höchsten Gewichtungsfaktor wird ein 182 mal höheres Gewicht zugesprochen als dem Fall mit dem niedrigsten Gewichtungsfaktor.

Analog zum Vorgehen im vorangegangenen Abschnitt wurden die beiden Regressionsmodelle auf der Grundlage der ungewichteten Daten, auf Basis einer reinen Ost-West-Gewichtung¹⁴ sowie auf Grundlage einer kombinierten Ost-West-/Repräsentativgewichtung geschätzt. Tabelle 4 zeigt die Ergebnisse. Vergleicht man zunächst die Befunde für die ungewichteten Daten mit der entsprechenden Spalte aus Tabelle 1, so ist klar zu erkennen, daß sich für die meisten Variablen (Bildung, Geschlecht, Alter, Wertorientierungen, Links-Rechts-Einstufung, PI) ein ähnliches Muster zeigt. Der Hauptunterschied zwischen Querschnitt- und Paneldaten besteht darin, daß die Bewertung der Performanz des politischen Systems im Panel keinen signifikanten Effekt hat. Zudem ist der Effekt der Parteiidentifikation erkennbar schwächer. Dabei ist allerdings zu beachten, daß auf Grund des wesentlich geringeren Stichprobenumfangs für die Paneldaten deutlich, d.h. ungefähr doppelt so große Standardfehler geschätzt werden.

Vergleicht man *innerhalb* der Paneldaten die Ergebnisse für die drei Gewichtungsszenarien, so zeigen sich erstaunlicherweise wiederum so gut wie keine substantiellen Unterschiede. Trotz der größeren Standardfehler weichen die Koeffizienten nur in einigen wenigen Fällen (Effekt der jüngsten Altersgruppe, Geschlecht, Wertorientierungen) um mehr als 0.2 Skaleneinheiten voneinander ab. Selbst in diesen Fällen bewegen sich die Diskrepanzen aber immer im Bereich von deutlich weniger als zwei Standardfehlern, sind also gemessen an der Präzision, mit der sich die Koeffizienten schätzen lassen, sehr klein. Ähnlich wie im Falle der Querschnittsdaten sind übrigens die Abweichungen zwischen ungewichteten und vollständig gewichteten Daten tendenziell kleiner als die Differenzen zwischen ungewichteten Daten und der reinen Ost-West-Gewichtung.

Der Vollständigkeit halber wurden auch aus diesem Modell die sozio-demographischen Kontrollvariablen entfernt. Wie im Querschnitt führt dies jedoch nicht zu einer Verstärkung der Unterschiede zwischen den drei

¹⁴ Da im Datensatz für die letzte Befragungswelle kein reines Ost-West-Gewicht bereitgestellt wird, wurde dies aus dem kombinierten Gewicht „cgges“ extrahiert.

Gewichtungsszenarien (Tabelle 5). Vielmehr kommt es tendenziell sogar zu einer Abschwächung der Unterschiede.¹⁵

Tabelle 6 zeigt schließlich die Ergebnisse für die Determinanten des Institutionenvertrauens. Wiederum ähneln bei wesentlich größeren Standardfehlern die Hauptergebnisse sehr stark den Befunden für den Querschnitt. Die soziodemographischen Variablen haben keinen nennenswerten Einfluß auf das Institutionenvertrauen, während eine negative Wahrnehmung der Wirtschaftslage und der allgemeinen Performanz des politischen Systems das Vertrauen in die Institutionen substantiell zu reduzieren vermögen. Zudem zeigt sich wiederum, daß die Anhänger der Regierungsparteien unabhängig davon den Institutionen einen Vertrauensbonus gewähren. Nennenswerte Abweichungen zwischen Panel und Querschnitt sind bei der Wirkung einzelner Einstellungsvariablen zu erkennen: Eine positive Bewertung des Nationalsozialismus hat im Panel einen etwas stärker negativen Effekt, während der analoge Effekt einer positiven Bewertung der DDR nicht mehr signifikant von null verschieden ist. Auch die (schwachen) Effekte der ideologischen Selbsteinstufung und der Wertorientierungen verschwinden im Panel gänzlich.

Die Differenzen, die sich *innerhalb* der Paneldaten durch die verschiedenen Gewichtungsszenarien ergeben, sind noch geringer als im Falle des Modells für die unkonventionelle Partizipation. In keinem Fall unterscheiden sich die Koeffizienten für gewichtete und ungewichtete Daten um mehr als 0.09 Skaleneinheiten. Die einzig inhaltlich interessante Abweichung betrifft die Wahrnehmung der persönlichen Wirtschaftslage. Diese hat im Falle der reinen Ost-West-Gewichtung wie im Querschnitt einen schwachen negativen Effekt, während dieser Effekt ohne Gewichtung bzw. unter einer kombinierten Ost-West-/Repräsentativgewichtung nicht signifikant von null verschieden ist. Selbst hier überlappen die Konfidenzintervalle der unter verschiedenen Gewichtungsszenarien geschätzten Koeffizienten aber so weit, daß es sich hier nicht um einen statistisch signifikanten Unterschied zwischen den Gewichtungsszenarien handeln kann. Selbst im Falle der hochgradig verzerrten Panelstichprobe sind deshalb alle Differenzen zwischen ungewichteten, ost-west-gewichteten und ost-west-/repräsentativgewichteten Daten letztlich vollkommen irrelevant.

4. Fazit

Ausgangspunkt dieses Beitrags war die Tatsache, daß Gewichtungsverfahren für viele Nutzer von sozialwissenschaftlichen Umfragedaten eine Art „black box“ darstellen, der sie mit einem gewissen Unbehagen gegenüberstehen. Während das Thema „Gewichtung“ in den gängigen Lehrbüchern kaum behandelt wird, wird in der Methodenausbildung häufig der Eindruck vermittelt, daß systematische Ausfälle nur die Randverteilungen betreffen, für die Analyse von Zusammenhängen von geringer Bedeutung sind und eine Gewichtung deshalb im Grunde entbehrlich ist. Wie in Abschnitt 2 auf Grundlage einer Reihe von Computersimulationen gezeigt werden konnte, ist der erste Teil dieser Aussage bedauerlicherweise falsch: Je nach

¹⁵ Die einzige Ausnahme ist hier die Wirkung der wahrgenommenen Performanz, wo sich die Differenzen zwischen der reinen Ost-West-Gewichtung einerseits und den ost-west-/repräsentativgewichteten bzw. ungewichteten Daten andererseits leicht verstärken.

Ausgestaltung des Ausfallmechanismus können systematische Ausfälle auch bei der Schätzung von Zusammenhängen erhebliche Verzerrungen bewirken. Diese Verzerrungen sind allerdings auch durch Gewichtungen kaum zu kompensieren.

Eine eingehende Analyse der Daten aus dem Forschungsprojekt hat überdies gezeigt, daß die Anwendung ambitionierter Gewichtungsprozeduren zumindest im Fall der beiden hier betrachteten Modelle keinerlei Einfluß auf die substantiellen Schlüsse hat, die aus den Befunden gezogen werden. Dies gilt selbst dann, wenn wie im Falle der Panelteilnehmer von 1994, 1998 und 2002 eine Stichprobe analysiert wird, die in extremer Weise von unit-nonresponse betroffen ist. Dies bedeutet leider nicht, daß unit-nonresponse unproblematisch wäre, sondern nur, daß ihre (normalerweise unbekannt) Folgen durch eine Gewichtung offensichtlich nicht zu kompensieren sind. Ein wichtiges Ziel jeder Befragung muß es deshalb sein, die Rate der (systematischen) Fälle von unit-nonresponse möglichst niedrig zu halten.

Umgekehrt sind Gewichtungen niemals unproblematisch. Vor allem im Falle von ex-post Gewichtungen, die systematische Antwortausfälle ausgleichen sollen, ist stets damit zu rechnen, daß die Varianz der Parameterschätzungen durch die Gewichtung steigt. Häufig wird dieser Verlust an Präzision die Vorteile einer weniger stark verzerrten Parameterschätzung überwiegen (Kish 1990: 127).

Im Ergebnis bestätigt dieser Beitrag deshalb die Empfehlung, die Winship/Radbill bereits 1994 ausgesprochen haben: Gewichtungen sind für Regressionsanalysen überhaupt nur dann „interessant“ (aber möglicherweise problematisch), wenn die Gewichte eine Funktion der abhängigen Variablen sind. Im Kontext der politischen Soziologie ist dies praktisch nie der Fall, da die Verteilung von Einstellungen und Handlungsabsichten in der Grundgesamtheit unbekannt ist und sich die Konstruktion von Gewichten normalerweise auf die obengenannten sozio-demographischen Variablen beschränkt, die ihrerseits kaum als abhängige Variablen analysiert werden.¹⁶ Sind die Gewichte hingegen eine Funktion der Variablen, die im Modell enthalten sind (was wiederum die Regel ist, da Sozialwissenschaftler aus guten Gründen gewohnheitsmäßig eine Reihe von sozio-demographischen Kontrollvariablen in ihre Modelle aufnehmen), so sollten bevorzugt die *ungewichteten* Daten analysiert werden, da die resultierenden Koeffizienten unverzerrt und konsistent sind und eine geringere Varianz aufweisen.

¹⁶ Eine mögliche Ausnahme stellt eine Gewichtung dar, die Verteilung der Wahlabsichten bzw. des berichteten Wahlverhalten an das tatsächliche Wahlergebnis anzugleichen versucht.

	ungewichtet	Ost/West	Ost/West*Rep
Geschlecht: männlich	0.27***	0.23***	0.29***
	(0.05)	(0.05)	(0.05)
Alter: <30	1.64***	1.63***	1.51***
	(0.09)	(0.09)	(0.10)
Alter: 30-45	1.55***	1.59***	1.46***
	(0.08)	(0.08)	(0.09)
Alter: 45-65	1.07***	1.06***	1.06***
	(0.07)	(0.08)	(0.09)
formale Bildung: mittel	0.50***	0.46***	0.53***
	(0.06)	(0.07)	(0.07)
formale Bildung: hoch	0.90***	0.86***	0.93***
	(0.07)	(0.07)	(0.07)
Region: Ost	-0.06	-0.08	-0.03
	(0.05)	(0.05)	(0.06)
LR-Position (zentriert)	0.16***	0.15***	0.14***
	(0.02)	(0.02)	(0.02)
Mischtyp	0.75***	0.81***	0.78***
	(0.06)	(0.07)	(0.07)
Postmaterialist	1.64***	1.73***	1.69***
	(0.08)	(0.08)	(0.09)
PI vorhanden	0.42***	0.33***	0.45***
	(0.06)	(0.06)	(0.07)
Performanz: unzufrieden	0.10***	0.07*	0.08**
	(0.03)	(0.03)	(0.03)
Konstante	3.17***	3.33***	3.20***
	(0.11)	(0.12)	(0.13)
N	8837	8837	8837
R2	0.20	0.20	0.19

Tabelle 1: Einfaches Partizipationsmodell (drei Gewichtungsvarianten)

	ungewichtet	Ost/West	Ost/West*Rep
LR-Position (zentriert)	0.19***	0.18***	0.17***
	(0.02)	(0.02)	(0.02)
Mischtyp	1.00***	1.05***	1.01***
	(0.06)	(0.07)	(0.07)
Postmaterialist	2.30***	2.39***	2.29***
	(0.08)	(0.08)	(0.09)
PI vorhanden	0.35***	0.23***	0.36***
	(0.06)	(0.06)	(0.07)
Performanz: unzufrieden	0.09**	0.04	0.06
	(0.03)	(0.03)	(0.03)
Konstante	4.57***	4.77***	4.71***
	(0.10)	(0.11)	(0.12)
N	8929	8929	8929
r2	0.11	0.11	0.10

Tabelle 2: Einfaches Partizipationsmodell ohne Kontrollvariablen (drei Gewichtungsvarianten)

	ungewichtet	Ost/West	Ost/West*Rep
Geschlecht: männlich	0.04*	0.04*	0.04*
	(0.02)	(0.02)	(0.02)
Alter: -30	-0.13***	-0.17***	-0.10**
	(0.03)	(0.03)	(0.04)
Alter: 30-45	-0.07*	-0.09**	-0.07*
	(0.03)	(0.03)	(0.03)
Alter: 45-65	0.01	0.01	-0.01
	(0.03)	(0.03)	(0.03)
formale Bildung: mittel	0.03	0.03	0.02
	(0.02)	(0.02)	(0.03)
formale Bildung: hoch	0.03	0.04	0.03
	(0.02)	(0.02)	(0.03)
Region: Ost	0.08***	0.05*	0.06*
	(0.02)	(0.02)	(0.02)
Regierungsanhänger	0.39***	0.37***	0.38***
	(0.02)	(0.02)	(0.02)
Allg. Wirtschaftslage schlecht	-0.11***	-0.10***	-0.11***
	(0.01)	(0.01)	(0.01)
Pers. Wirtschaftslage schlecht	-0.08***	-0.09***	-0.08***
	(0.01)	(0.01)	(0.01)
Performanz: unzufrieden	-0.26***	-0.25***	-0.26***
	(0.01)	(0.01)	(0.01)
Mischtyp	-0.03	-0.03	-0.03
	(0.02)	(0.02)	(0.02)
Postmaterialist	-0.11***	-0.09**	-0.12***
	(0.03)	(0.03)	(0.03)
LR-Position (zentriert)	-0.03***	-0.02***	-0.03***
	(0.01)	(0.01)	(0.01)
NS: gute Seiten	-0.03***	-0.03**	-0.02
	(0.01)	(0.01)	(0.01)
DDR: gute Seiten	-0.07***	-0.05***	-0.06***
	(0.01)	(0.01)	(0.01)
Konstante	4.81***	4.74***	4.81***
	(0.05)	(0.06)	(0.06)
N	7018	7018	7018
r ²	0.28	0.26	0.27

Tabelle 3: Determinanten des Institutionenvertrauens

	ungewichtet	Ost/West	Ost/West*Rep
Geschlecht: männlich	0.36**	0.13	0.36*
	(0.12)	(0.15)	(0.16)
Alter: -30	1.02**	1.25***	0.75*
	(0.32)	(0.32)	(0.37)
Alter: 30-45	1.12***	1.06***	1.06***
	(0.18)	(0.23)	(0.22)
Alter: 45-65	0.73***	0.63**	0.59**
	(0.15)	(0.21)	(0.20)
formale Bildung: mittel	0.73***	0.60**	0.70**
	(0.16)	(0.20)	(0.21)
formale Bildung: hoch	1.05***	0.93***	1.07***
	(0.16)	(0.19)	(0.21)
Region: Ost	0.11	0.15	0.11
	(0.13)	(0.14)	(0.16)
LR-Position (zentriert)	0.12**	0.14*	0.12*
	(0.04)	(0.05)	(0.05)
Mischtyp	0.66***	0.46*	0.69***
	(0.14)	(0.18)	(0.20)
Postmaterialist	1.09***	1.01***	1.34***
	(0.19)	(0.22)	(0.26)
PI vorhanden	0.15	0.12	-0.03
	(0.16)	(0.20)	(0.21)
Performanz: unzufrieden	-0.07	-0.16	-0.07
	(0.07)	(0.08)	(0.09)
Konstante	4.28***	4.91***	4.49***
	(0.29)	(0.38)	(0.38)
N	1297	1297	1297
r2	0.17	0.16	0.19

Tabelle 4: Einfaches Partizipationsmodell (drei Gewichtungsvarianten, Panel)

	ungewichtet	Ost/West	Ost/West*Rep
LR-Position (zentriert)	0.15***	0.15**	0.14*
	(0.04)	(0.05)	(0.06)
Mischtyp	1.00***	0.81***	1.01***
	(0.14)	(0.18)	(0.19)
Postmaterialist	1.72***	1.63***	1.95***
	(0.19)	(0.21)	(0.24)
PI vorhanden	0.02	-0.09	-0.08
	(0.16)	(0.20)	(0.22)
Performanz: unzufrieden	-0.08	-0.20*	-0.04
	(0.07)	(0.08)	(0.09)
Konstante	5.61***	6.12***	5.70***
	(0.27)	(0.34)	(0.36)
N	1298	1298	1298
r2	0.08	0.09	0.11

Tabelle 5: Einfaches Partizipationsmodell ohne Kontrollvariablen (drei Gewichtungsvarianten, Panel)

	ungewichtet	Ost/West	Ost/West*Rep
--	-------------	----------	--------------

Geschlecht: männlich	0.08*	0.15**	0.05
	(0.04)	(0.05)	(0.05)
Alter: <30	-0.12	-0.15	-0.18
	(0.11)	(0.11)	(0.11)
Alter: 30-45	-0.12	-0.13	-0.17*
	(0.06)	(0.07)	(0.07)
Alter: 45-65	-0.00	-0.01	-0.05
	(0.05)	(0.06)	(0.06)
formale Bildung: mittel	-0.09	-0.04	-0.06
	(0.05)	(0.06)	(0.07)
formale Bildung: hoch	-0.09	-0.09	-0.03
	(0.05)	(0.06)	(0.07)
Region: Ost	0.04	0.03	0.03
	(0.05)	(0.05)	(0.06)
Regierungsanhänger	0.31***	0.35***	0.28***
	(0.04)	(0.05)	(0.05)
Allg. Wirtschaftslage schlecht	-0.10***	-0.10**	-0.11**
	(0.03)	(0.03)	(0.03)
Pers. Wirtschaftslage schlecht	-0.01	-0.08**	0.03
	(0.03)	(0.03)	(0.04)
Performanz: unzufrieden	-0.27***	-0.25***	-0.23***
	(0.02)	(0.03)	(0.03)
Mischtyp	-0.01	-0.05	-0.01
	(0.05)	(0.06)	(0.06)
Postmaterialist	0.08	0.05	0.16*
	(0.06)	(0.07)	(0.08)
LR-Position (zentriert)	-0.01	-0.01	-0.02
	(0.01)	(0.02)	(0.02)
NS: gute Seiten	-0.07***	-0.08**	-0.09***
	(0.02)	(0.03)	(0.03)
DDR: gute Seiten	-0.01	-0.00	-0.01
	(0.02)	(0.02)	(0.03)
Konstante	4.67***	4.76***	4.52***
	(0.13)	(0.15)	(0.17)
N	1144	1144	1144
r ²	0.26	0.31	0.24

Tabelle 6: Determinanten des Institutionenvertrauens

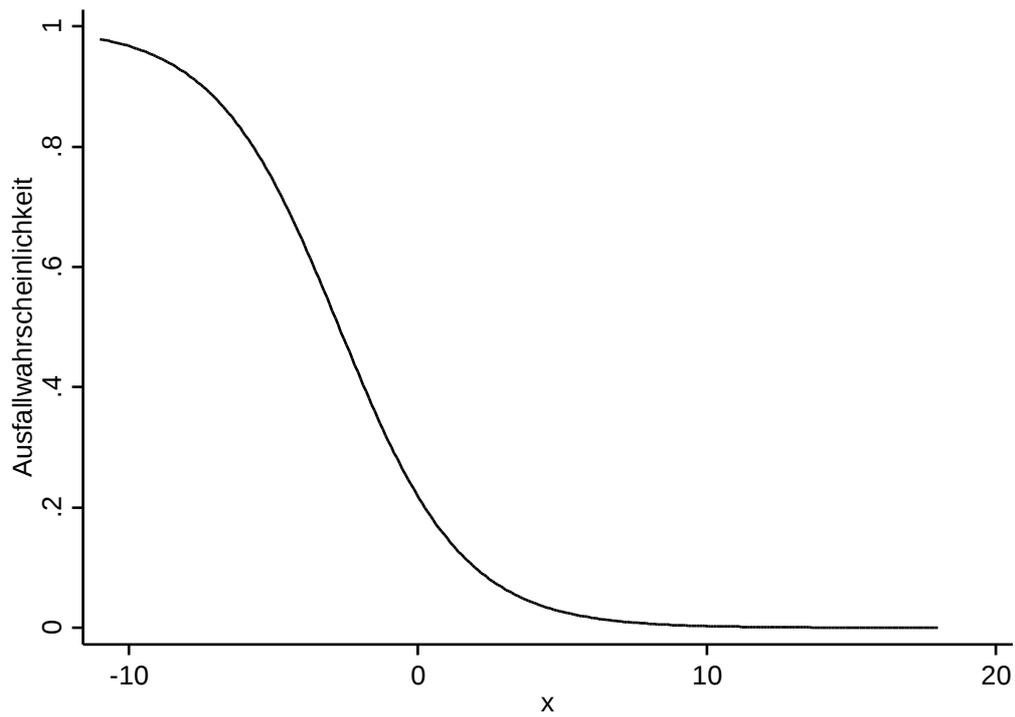


Abbildung 1: Wahrscheinlichkeit einer unit-nonresponse in Abhängigkeit von x

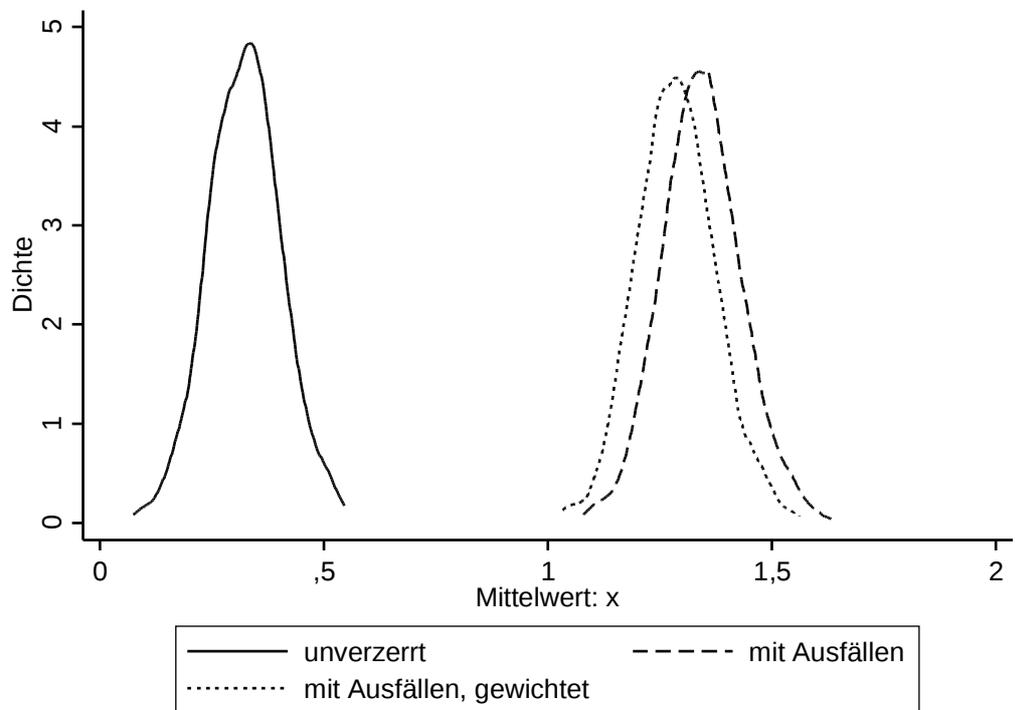


Abbildung 2: Mittelwert von x ohne Ausfälle, mit Ausfällen und mit Ausfällen/Gewichtung

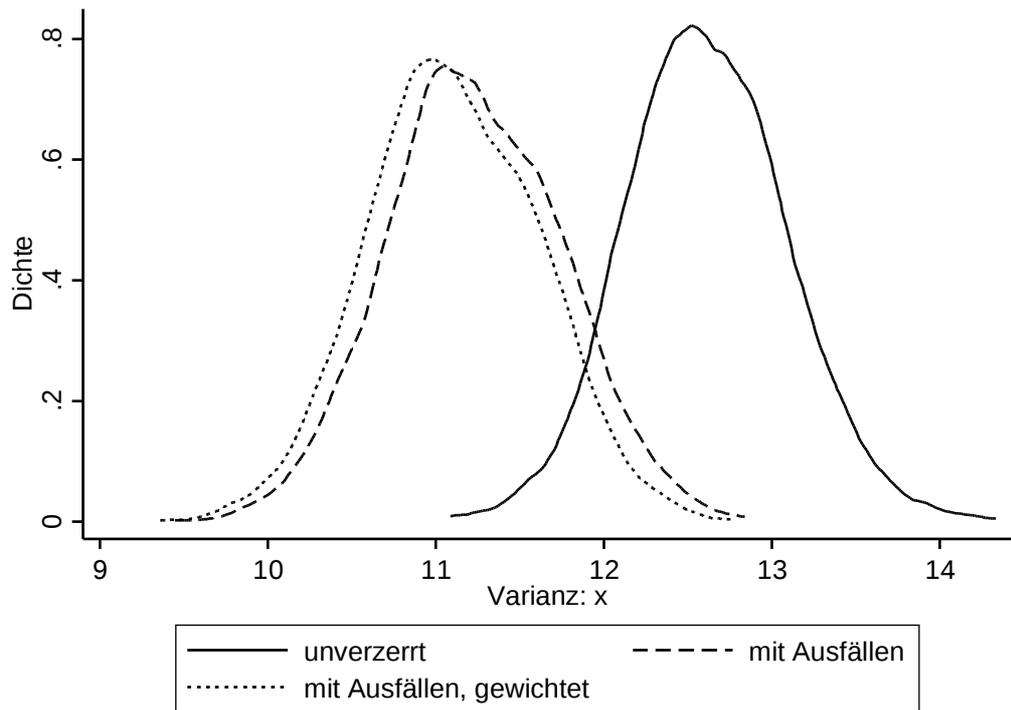


Abbildung 3: Varianz von x ohne Ausfälle, mit Ausfällen und mit Ausfällen/Gewichtung

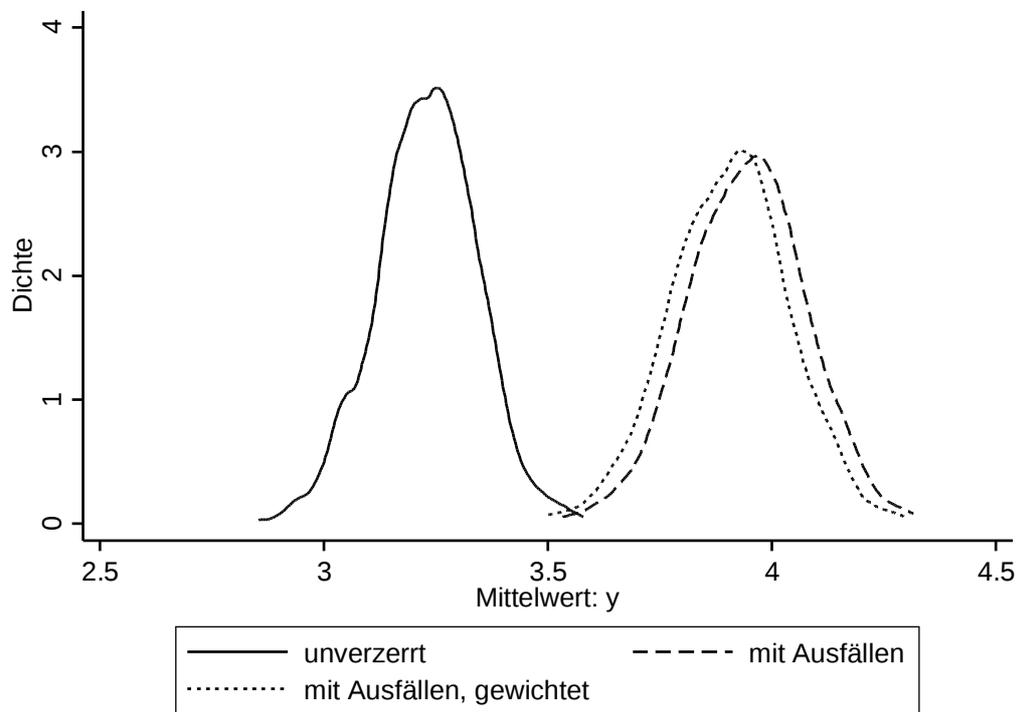


Abbildung 4: Mittelwert von y ohne Ausfälle, mit Ausfällen und mit Ausfällen/Gewichtung

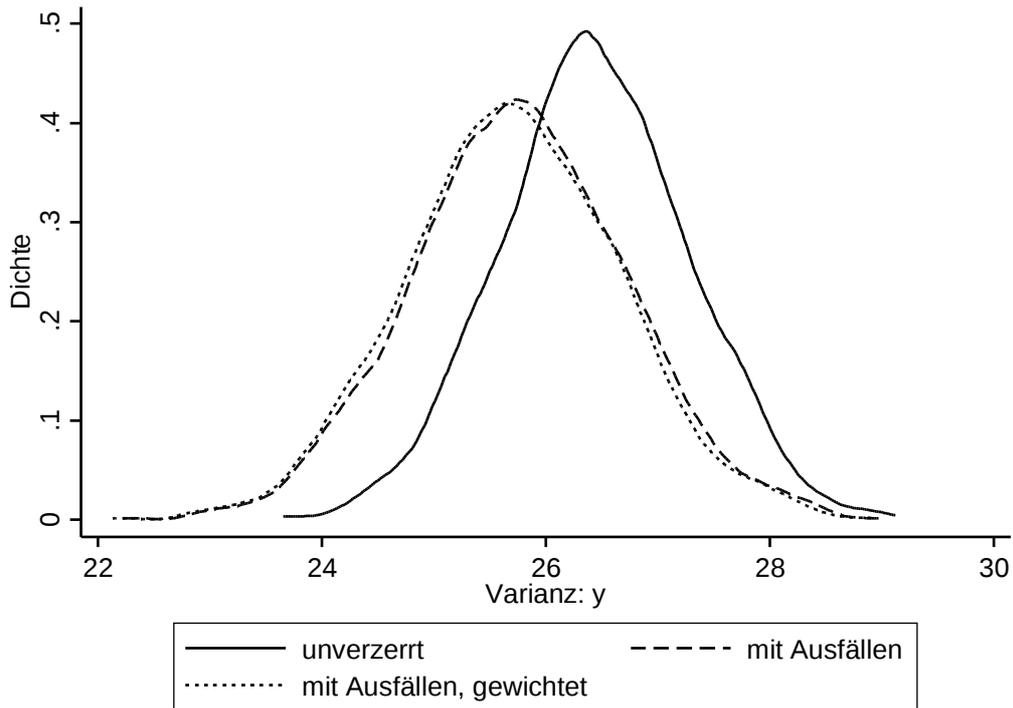


Abbildung 5: Varianz von y ohne Ausfälle, mit Ausfällen und mit Ausfällen/Gewichtung

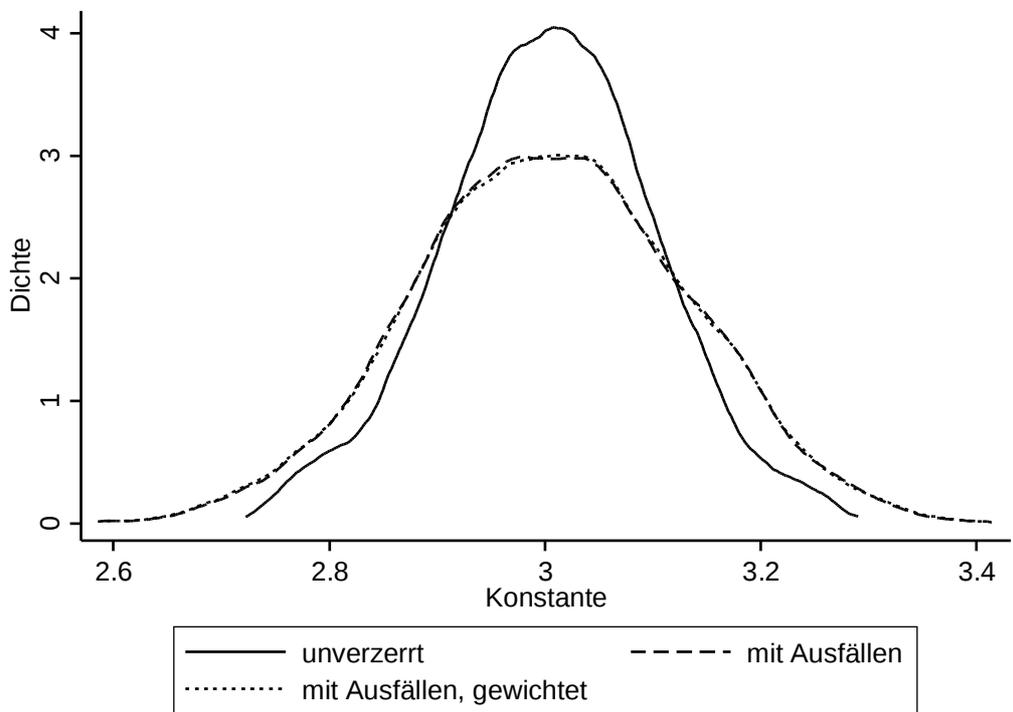


Abbildung 6: Verteilung der Konstante ohne Ausfälle, mit Ausfällen und mit Ausfällen/Gewichtung

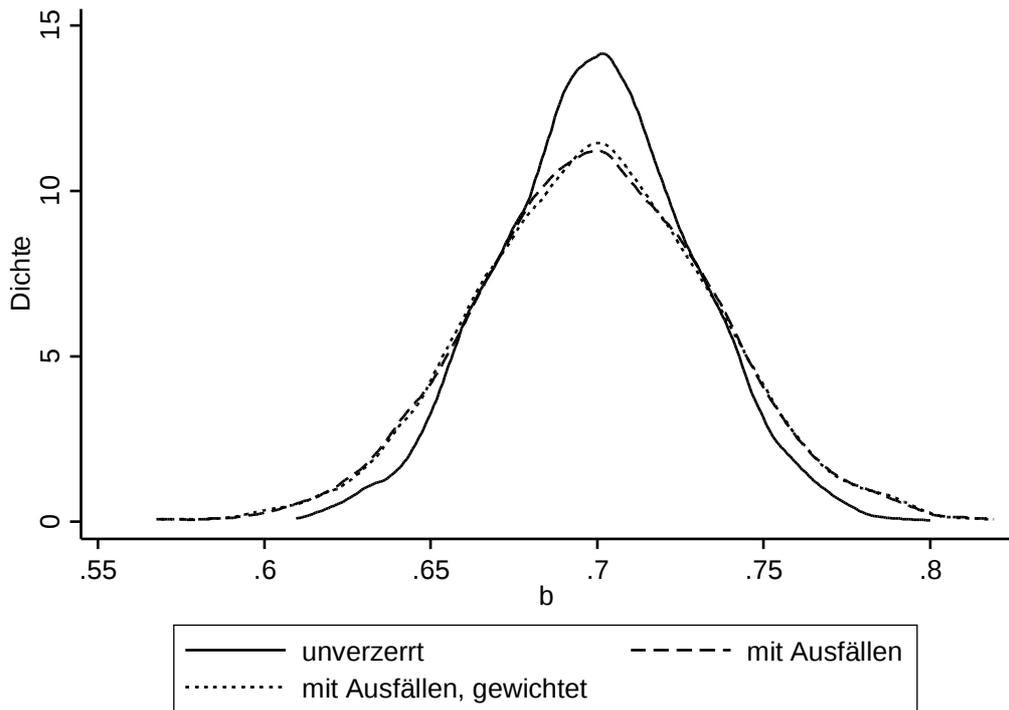


Abbildung 7: Verteilung von b ohne Ausfälle, mit Ausfällen und mit Ausfällen/Gewichtung

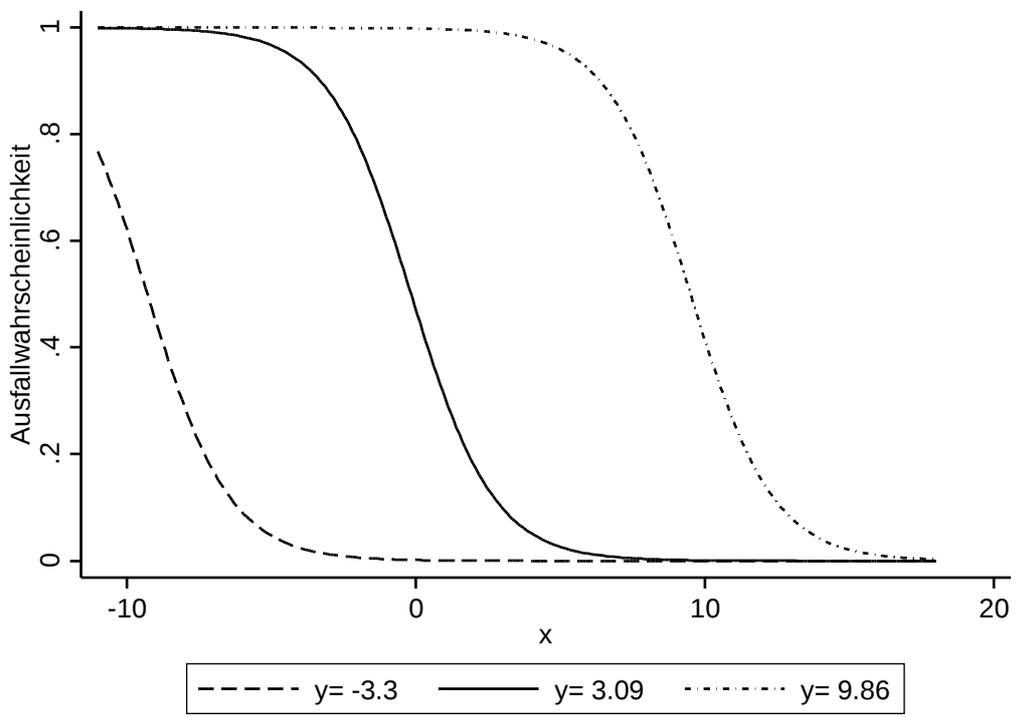


Abbildung 8: Wahrscheinlichkeit einer unit-nonresponse in Abhängigkeit von x und y (Mechanismus 2)

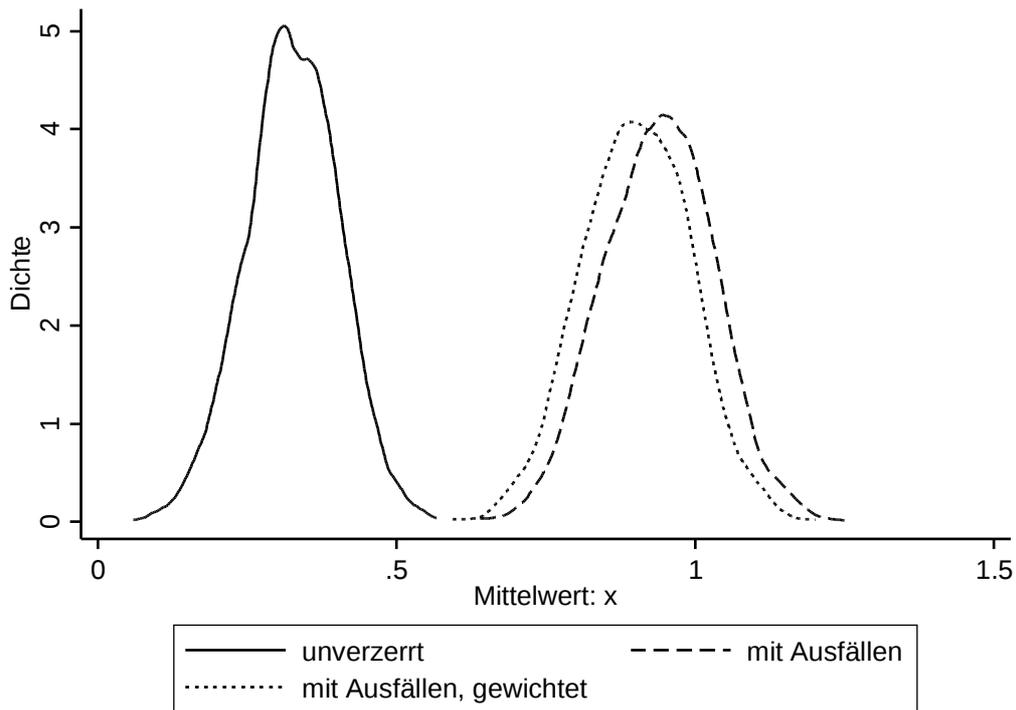


Abbildung 9: Mittelwert von x ohne Ausfälle, mit Ausfällen und mit Ausfällen/Gewichtung (Mechanismus 2)

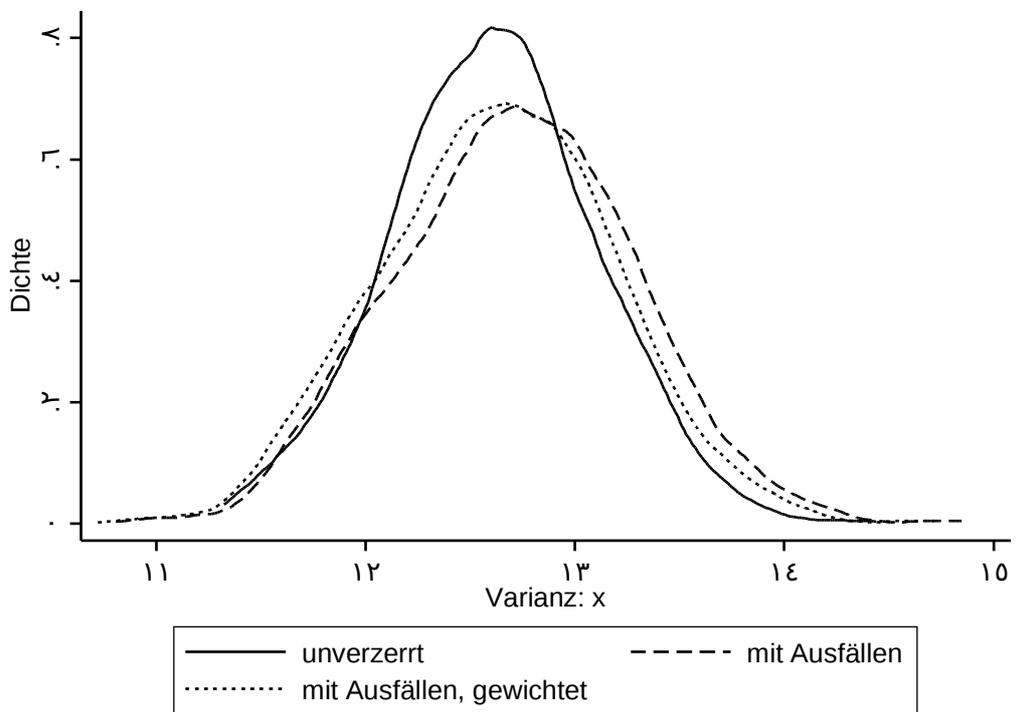


Abbildung 10: Varianz von x ohne Ausfälle, mit Ausfällen und mit Ausfällen/Gewichtung (Mechanismus 2)

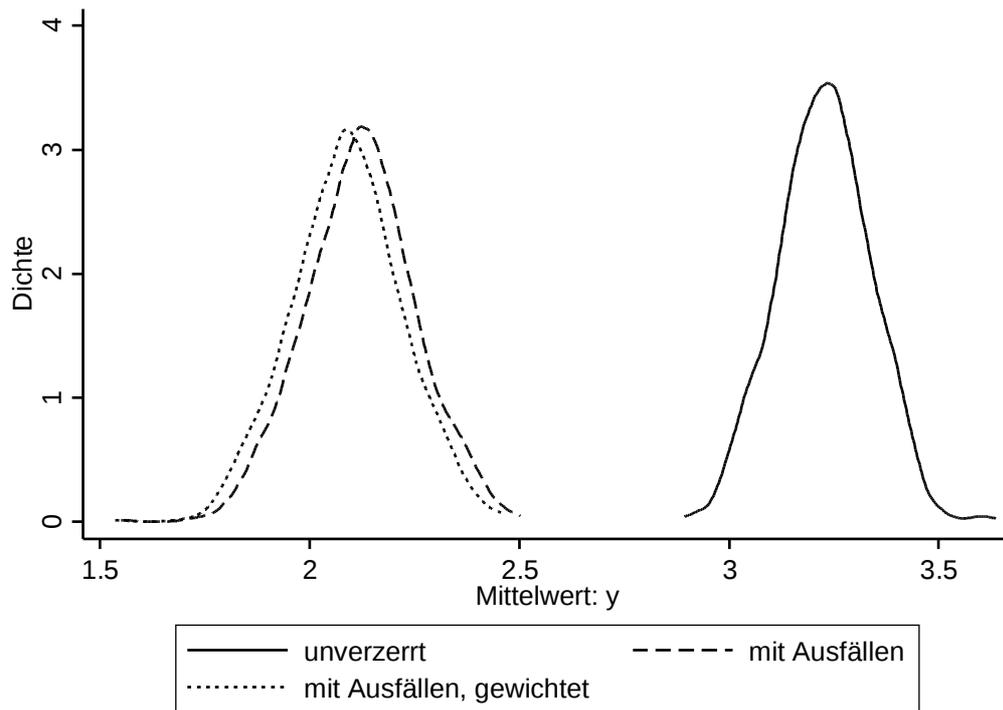


Abbildung 11: Mittelwert von y ohne Ausfälle, mit Ausfällen und mit Ausfällen/Gewichtung (Mechanismus 2)

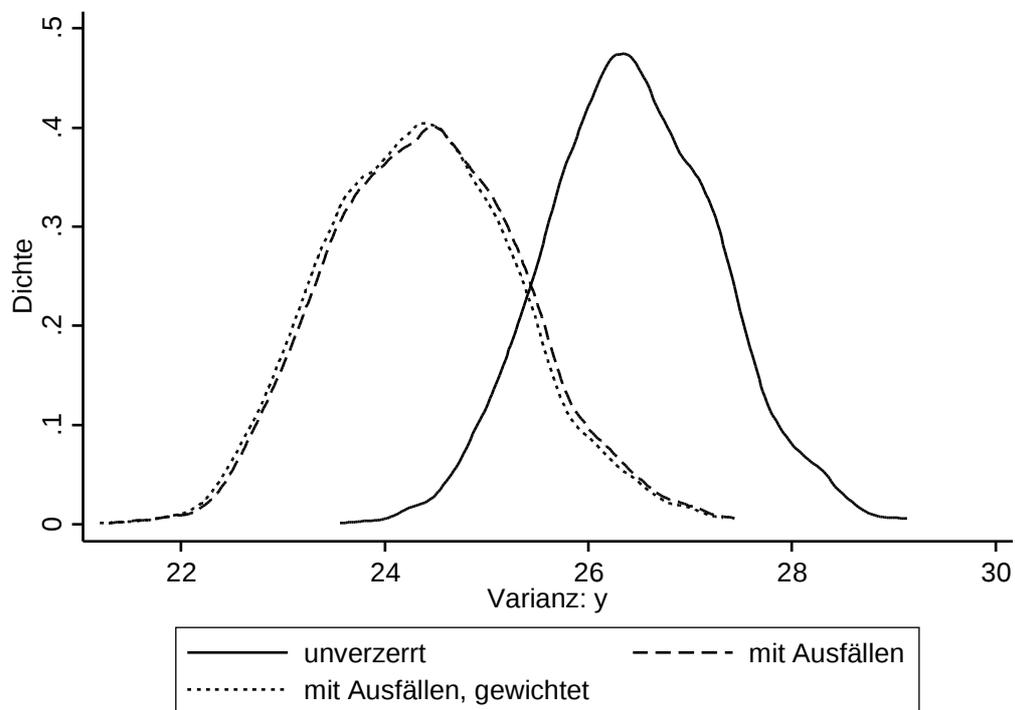


Abbildung 12: Varianz von y ohne Ausfälle, mit Ausfällen und mit Ausfällen/Gewichtung (Mechanismus 2)

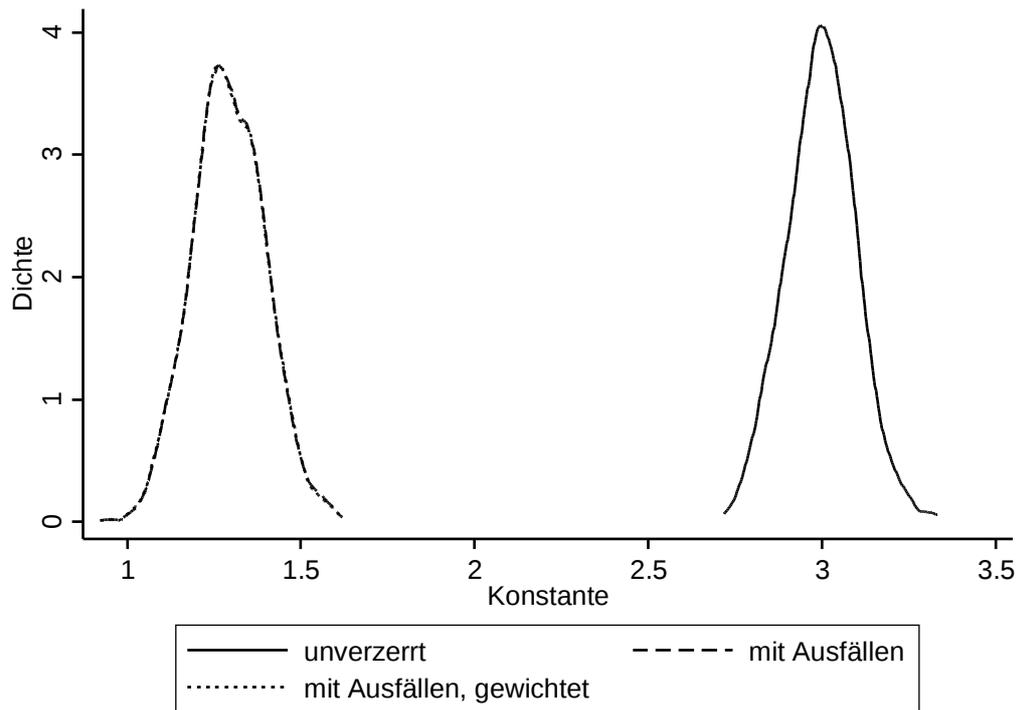


Abbildung 13: Verteilung der Konstante ohne Ausfälle, mit Ausfällen und mit Ausfällen/Gewichtung (Mechanismus 2)

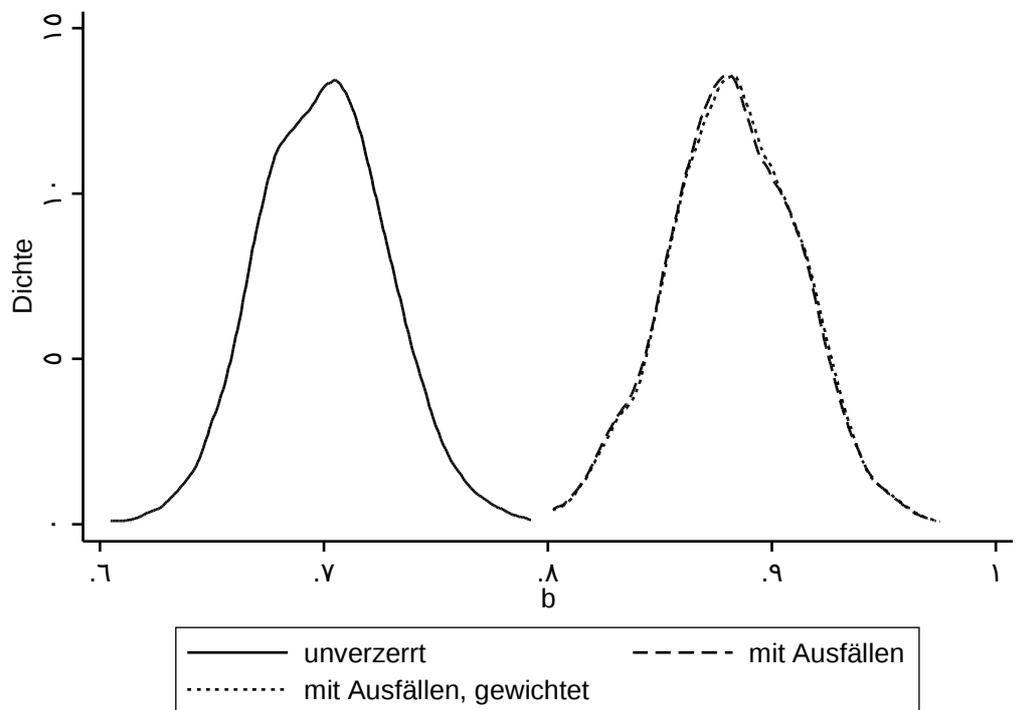


Abbildung 14: Verteilung von b ohne Ausfälle, mit Ausfällen und mit Ausfällen/Gewichtung (Mechanismus 2)

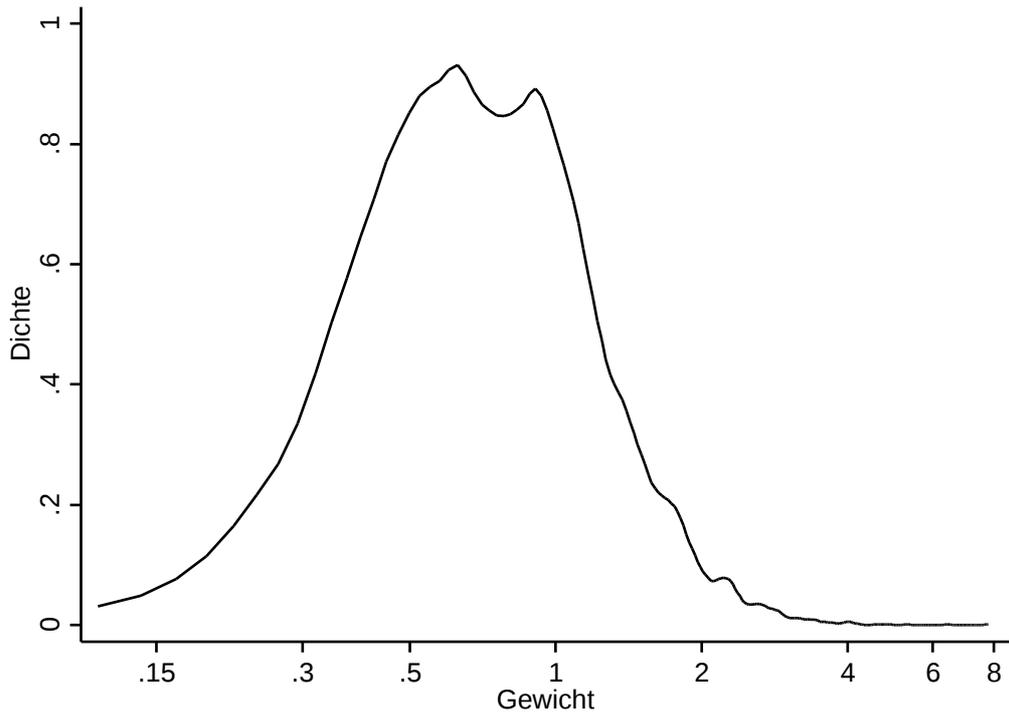


Abbildung 15: Verteilung der Gewichtungvariable „vgges“ im Querschnitt (x-Achse logarithmisch skaliert)

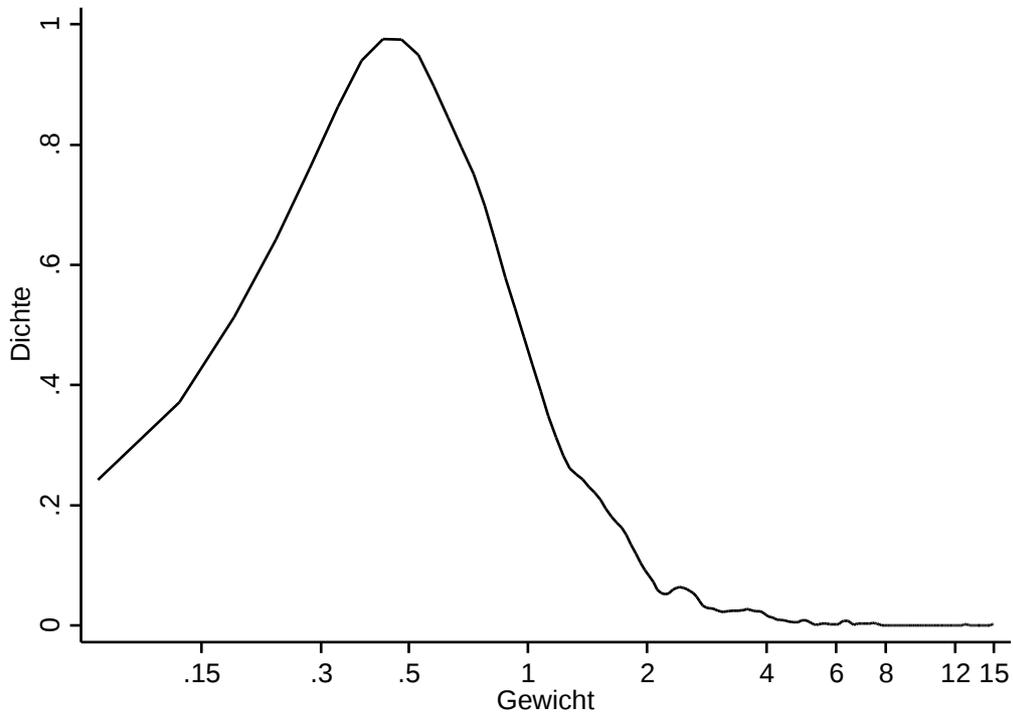


Abbildung 16: Verteilung der Gewichtungvariable „vgges“ im Panel (x-Achse logarithmisch skaliert)