

Gruppenvergleiche bei Regressionen mit binären abhängigen Variablen – Probleme und Fehleinschätzungen am Beispiel von Bildungschancen im Kohortenverlauf

Group Comparisons for Regression Models with Binary Dependent Variables – Problems and Pitfalls Illustrated by Differences in Educational Opportunities between Cohorts

Katrin Auspurg und Thomas Hinz*

Universität Konstanz, Fachbereich Geschichte und Soziologie, Universitätsstr. 10, 78457 Konstanz, Germany
katrin.auspurg@uni-konstanz.de; thomas.hinz@uni-konstanz.de

Zusammenfassung: Die vorliegende Forschungsnotiz verweist auf eine bekannte, aber selten beachtete Problematik bei Vergleichen von Koeffizienten aus Regressionen mit binären abhängigen Variablen zwischen Gruppen. Damit Vergleiche von Logit- und Probit-Koeffizienten sowie Odds-Ratios (OR) über Gruppen oder Kohorten hinweg tragfähig sind, muss angenommen werden, dass die unbeobachtete Heterogenität in allen betrachteten Gruppen gleich ist. Dies ist zumindest bei Vergleichen von Kohorten oder Schätzungen, die Datensätze aus unterschiedlichen Ländern oder Erhebungsdesigns nutzen, eine sehr unrealistische Annahme. Wir schlagen daher vor, statt der OR die durchschnittlichen Marginaleffekte für Gruppenvergleiche heranzuziehen. Weiterhin verweisen wir auf eine Möglichkeit, Gruppenunterschiede in solchen Modellen auf statistische Signifikanz zu prüfen. Anhand des Beispiels von Bildungschancen im Kohortenvergleich lässt sich veranschaulichen, dass bei Beachtung der Problematik unbeobachteter Heterogenität wichtige Schlussfolgerungen anders ausfallen.

Schlagerworte: Binäre Regressionsmodelle; Logistische Regression; Gruppenvergleiche; Unbeobachtete Heterogenität; Bildungssoziologie; SOEP.

Summary: This research note refers to a known, but rarely noticed problem which arises when coefficients from regression models with binary dependent variables are compared over groups like cohorts or social classes. In order to attain valid and viable comparisons of coefficients and odds ratios (OR) from logit and probit models between groups, it has to be assumed that the unobserved heterogeneity is equal for all these groups. This is obviously an unrealistic assumption if data stem from different cohorts or if estimations are based on data from different countries and samples. Therefore, we propose for group comparisons the use of average marginal effects instead of OR. Moreover, we suggest a method of testing group differences in such models for statistical significance. Using the example of comparing educational opportunities over different birth cohorts, we illustrate that considering the problem of unobserved heterogeneity leads to significantly different conclusions.

Keywords: Binary Regression Models; Logistic Regressions; Group Comparisons; Unobserved Heterogeneity; Sociology of Education; SOEP.

1. Einleitung

Die logistische Regression mit einer dichotomen abhängigen Variablen gehört zu den Routineverfahren in den Sozialwissenschaften. Anwendungsbeispiele sind die Analyse von Eintrittswahrscheinlichkeiten

binärer Zustände (z. B. Betroffenheit von Arbeitslosigkeit, Bestehen einer Partnerschaft) oder von Entscheidungen zwischen binären Alternativen (z. B. Einschlagen eines höheren Bildungsweges, Teilnahme an Protesten). In der Soziologie sind bei solchen Fragestellungen in aller Regel Gruppen- oder Kohortenvergleiche von hohem Interesse: Führen etwa bei In- und Ausländern dieselben Risikofaktoren zur Arbeitslosigkeit? Hat die soziale Herkunft in verschiedenen Geburtskohorten einen gleichermaßen starken Einfluss auf die Bildungsnachfrage oder ist es gelungen, schichtspezifische Bildungsungleichheiten abzubauen?

* Für die sehr sorgfältige Durchsicht unseres Manuskripts bedanken wir uns bei einem anonymen Gutachter. Sehr hilfreiche Hinweise verdanken wir zudem Carina Mood, Winfried Pohlmeier, Ben Jann und Maarten Buis. Für mögliche Fehler oder verkürzte Darstellungsweisen sind allein die Autoren verantwortlich.

Neben der logistischen Regression werden seltener auch andere nichtlineare Modelle für dichotome abhängige Variablen herangezogen, etwa Probit-Regressionen. Alle nichtlinearen Modelle für solche Variablen werden im Folgenden als *binäre* Regressionen bezeichnet. Mit ihrer Hilfe werden relative Chancenverhältnisse (bzw. *Odds Ratios*, kurz OR) für verschiedene Gruppen (z. B. soziale Klassen, Länder, Männer versus Frauen) oder Kohorten ermittelt und dann in Form von Tabellen oder grafischen Darstellungen miteinander verglichen. Die Interpretation der Chancenverhältnisse entspricht durchaus dem *State of the Art*, sind diese Maße doch bekanntlich weitgehend unempfindlich gegenüber zeitlichen Veränderungen von allgemeinen (Bildungs-)Gelegenheiten oder Randverteilungen (wie denen sozialer Klassen).¹ Dennoch weist die Methode einige Tücken auf, gerade wenn die Ergebnisse für verschiedene Gruppen, Kohorten oder Stichproben (Länder) verglichen werden. Über die dabei bestehende Gefahr von Fehlschlüssen und Abhilfen wurde zwar insbesondere in der Ökonometrie (z. B. Wooldridge 2002) viel publiziert. Bislang ist dieses Wissen jedoch nicht in die Standardlehrbücher der sozialwissenschaftlichen Datenanalyse eingeflossen. Gerade in der deutschsprachigen Soziologie ignorieren viele empirisch arbeitenden Sozialwissenschaftler weiterhin die Probleme. Dies mag an der oftmals sehr technischen Darstellung der zugrunde liegenden Statistik liegen.

Ziel der vorliegenden Forschungsnotiz ist es daher, in möglichst eingängiger Weise über die Tragfähigkeit von Gruppenvergleichen bei binären Regressionen zu informieren. Es soll lediglich auf solche Statistikkennnisse aufgebaut werden, die dem durchschnittlich methodisch vorgebildeten Leser vertraut sein dürften. Konkretes Anschauungsmaterial bietet der kürzlich von Andreas Hadjar und Joël Berger (2010) in dieser Zeitschrift erschienene Beitrag „Dauerhafte Bildungsungleichheiten in Westdeutschland, Ostdeutschland und der Schweiz: Eine Kohortenbetrachtung der Ungleichheitsdimensionen soziale Herkunft und Geschlecht“. Die dort vorgelegten Analysen fokussieren ein klassisches Thema der Bildungssoziologie: Inwieweit hat die Bildungsexpansion zum Abbau herkunfts- und geschlechtsspezifischer Ungleichheiten geführt, und zwar in länder- und kohortenspezifischer Weise? Mit diesen Fragen wird an eine lange Tradition in

der Bildungssoziologie angeknüpft.² Hadjar und Berger bewegen sich also – wie sie selbst anmerken (vgl. Hadjar & Berger 2010: 183) – inhaltlich und empirisch auf gut erschlossenem Terrain. Auch die eingesetzten statistischen Methoden folgen einem weithin praktizierten Standard. Gleichwohl werden die folgenden Ausführungen und beispielhaften Berechnungen zeigen, dass alternative Methoden zu besser abgesicherten Schätzungen und anderen Schlussfolgerungen führen.

2. Das Problem unbeobachteter Heterogenität bei logistischen Regressionen

Eine häufig nicht beachtete Besonderheit von binären Regressionen besteht darin, dass bei ihnen, anders als bei linearen *Ordinary Least Square*- (OLS-) Regressionen, die Regressionskoeffizienten mit der Residualvarianz bzw. unbeobachteten Heterogenität konfundiert sind – und dies selbst dann, wenn die im Modell nicht berücksichtigten erklärenden Variablen mit den im Modell enthaltenen erklärenden Variablen unkorreliert sind (vgl. Allison 1999; Mood 2010; Wooldridge 2002). In der Konsequenz können die Regressionskoeffizienten zwischen Gruppen oder Stichproben unterschiedlich ausfallen, auch wenn die kausalen Einflüsse der betreffenden Variablen übereinstimmen. Dafür genügt es, dass sich die Residualvarianz zwischen den Gruppen bzw. Stichproben unterscheidet. Nochmals anders gesagt: Liegen gruppenspezifischen Residualvarianzen vor (*Heteroskedastizität*), werden die Koeffizienten bei binären Regressionen – im Gegensatz zu OLS-Regressionen – nicht mehr unverzerrt (*konsistent*) geschätzt.

Bevor die hiermit verbundene Gefahr inhaltlicher Fehlschlüsse an einem Beispiel aufgezeigt wird, wollen wir im Folgenden kurz die statistischen Grundlagen erläutern. Besonders verständlich lassen sich diese mittels der Motivation vermitteln, binäre Regressionen als so genannte *Schwellenwertmodelle* aufzufassen (z. B. Allison 1999; Brüderl 2000; Mood 2010). Binäre Regressionsmodelle schätzen bekanntlich die Wahrscheinlichkeit P , dass ein bestimmtes Ereignis vorliegt – in unserem Anwendungsbeispiel ist es das Vorliegen eines höheren Bildungsabschlusses bei den Befragten i ($i = 1, \dots, n$). Die gängige Notation für diese zu schätzende

² So finden sich bei Blossfeld 1993, Handl 1985, Henz & Maas 1995 und Schimpl-Neimanns 2000 ähnliche Kohortenuntersuchungen; vgl. zudem Arum et al. 2007 sowie Blossfeld & Shavit 1993 für komparative Analysen von Bildungssystemen.

¹ Zur diesbezüglichen Methodendiskussion in der Bildungssoziologie vgl. Handl 1985 sowie Müller & Haun 1994.

Wahrscheinlichkeit lautet P_i ($y_i = 1$), wobei y_i eine Dummy-Variable darstellt, die den Wert 1 einnimmt, wenn das interessierende Ereignis (der Bildungsabschluss) vorliegt, und sonst den Wert 0 erhält. Die zugrunde liegende Vorstellung für die Formulierung eines Schwellenwertmodells ist nun, dass dieses Ereignis immer dann eintritt, wenn eine latente, nicht-beobachtbare Variable y_i^* eine bestimmte Schwelle überschreitet. Im vorliegenden Anwendungsfall kann man sich unter y_i^* etwa den individuellen Bildungsnutzen oder die Bildungsmotivation der Befragten vorstellen. Da die Wahl der Schwelle für die Schätzergebnisse der Regressionskoeffizienten unerheblich ist,³ wird sie konventionell auf den Wert 0 gesetzt, formal:

$$y = 1 \text{ wenn } y^* > 0 \\ y = 0 \text{ wenn } y^* \leq 0$$

Des Weiteren wird davon ausgegangen, dass diese latente Variable linear von bis zu p verschiedenen Kovariaten x_j abhängt ($j = 1, \dots, p$). Hinzu kommt eine Fehlerkomponente (*Residuum*), die der Vorlage von Allison (1999) folgend – und dies ist nun ein Aspekt, der von den gängigen Darstellungen in soziologischen Einführungsbüchern abweicht – als $\sigma_g \varepsilon_i$ notiert wird, um die Möglichkeit von Variationen im Ausmaß unbeobachteter Heterogenität bzw. Heteroskedastizität zwischen einzelnen Gruppen (die hier mit dem Index g kenntlich gemacht werden; $g = 1, \dots, o$) explizit zuzulassen. Je größer der Betrag des Parameters σ_g ist, umso größer ist die unsystematische Fehlerkomponente bzw. unbeobachtete Heterogenität in der Gruppe g . Insgesamt ergibt sich somit folgendes Modell für die latente Variable y_{ig}^* in Gruppe g :

$$y_{ig}^* = \alpha_{0g} + \alpha_{1g} x_{i1g} + \dots + \alpha_{pg} x_{ipg} + \sigma_g \varepsilon_i \quad (1)$$

wobei α_{0g} für die Konstante steht und α_{1g} bis α_{pg} die Regressionskoeffizienten in diesem linearen Modell bezeichnen. Unterstellt man für den Fehlerterm ε_i eine logistische Verteilung,⁴ resultiert daraus das bekannte logistische Regressionsmodell:

$$\text{logit}_{ig} = \log \left(\frac{P_{ig}}{1 - P_{ig}} \right) = \beta_{0g} + \beta_{1g} x_{i1g} + \dots + \beta_{pg} x_{ipg} \quad (2)$$

Weitgehend unbekannt ist, dass die Regressionskoeffizienten β in dieser Gleichung mit den Koeffizienten α des linearen Modells in folgender Beziehung stehen (vgl. Allison 1999: 189):

$$\beta_{ig} = \frac{\alpha_{ig}}{\sigma_g} \quad (3)$$

Dieser Zusammenhang gilt ebenso für andere binäre Regressionsmodelle (wie beispielsweise Probit-Modelle). Da *per se* nicht zugleich β_{ig} und σ_g identifizierbar sind (das Regressionsmodell ist in dieser Hinsicht unterdeterminiert), wird die unbeobachtete Heterogenität σ_g üblicherweise auf den Wert $\frac{\pi}{\sqrt{3}}$ normiert – womit β_{ig} einfach mit α_{ig} , dividiert durch eine Konstante, gleichgesetzt wird.⁵ Diese Vereinfachung ist vermutlich auch der Grund, weshalb der Zusammenhang aus Gleichung (3) in vielen Darstellungen vernachlässigt wird. Für die üblichen Tests auf die Signifikanz des Einflusses der Merkmale x_{ig} ist diese Konvention auch unerheblich, denn $\beta_{ig} = 0$ impliziert stets auch, dass $\alpha_{ig} = 0$; das heißt, die signifikante Abweichung vom Wert Null kann unabhängig von σ_g geprüft werden (Allison 1999: 189; Brüderl 2000). Für das Problem unbeobachteter Heterogenität ist der Zusammenhang aus Gleichung (3) hingegen von zentraler Bedeutung. Denn mit der analogen Festlegung der Varianz in mehreren zu vergleichenden Gruppen (oder Stichproben) wird impliziert, dass das Ausmaß unbeobachteter Heterogenität in den einzelnen Gruppen oder Stichproben identisch ist (also etwa bei einem Vergleich der Gruppen k und l : $\sigma_k = \sigma_l$). Dies dürfte in den meisten Fällen eine sehr unrealistische Annahme sein. Sicher ist diese Bedingung für die im Folgenden noch näher dargestellte empirische Untersuchung von Bildungsungleichheiten, bei der nur drei verschiedene erklärende Variablen (Geschlecht,

³ Der Grund dafür ist, dass der Schwellenwert allein die Konstante α_0 des nachfolgend dargestellten Regressionsmodells beeinflusst, nicht aber die eigentlich interessierenden Regressionskoeffizienten α_1 bis α_p (Brüderl 2000).

⁴ Die logistische Verteilung ist eine gängige Funktion, um Wachstumsprozesse mit Sättigungsphänomenen zu modellieren. Sie ist definiert als: $F(x) = P(X \leq x) =$

$$\frac{1}{1 + \exp \left(-\frac{(x - \mu) \pi}{\sigma \sqrt{3}} \right)}$$

wobei $\exp(\cdot)$ die Exponentialfunktion, μ den Erwartungswert von x und π die Kreiszahl (3,14...) notiert.

⁵ Der Grund für die Normierung auf diesen Wert ist, dass sich hiermit vereinfachte Formeln für die logistische Verteilung ergeben. Für σ_g ist folgender Zusammenhang mit einer nicht in Gleichung (1) berücksichtigten, normalverteilten Kovariate v_g mit Varianz τ_g^2 bekannt: $\sigma_g^2 = \lambda_g^2 \tau_g^2 + 1$, wobei λ_g für den zu v_g gehörenden Regressionskoeffizienten steht ($y_g^* = \alpha_{0g} + \alpha_{1g} x_{1g} + \dots + \alpha_{pg} x_{pg} + \lambda_g v_g$; vgl. Wooldridge 2002: 470). Nur in dem sehr unwahrscheinlichen Fall, dass der Einfluss aller nicht berücksichtigten erklärenden Variablen auf y_g^* nicht zwischen den Gruppen variiert und diese unberücksichtigten Variablen überdies in allen Gruppen die gleiche Varianz τ aufweisen, unterscheidet sich die unbeobachtete Heterogenität σ nicht zwischen den Gruppen.

Kohorte und soziale Herkunft) herangezogen werden, eine sehr starke Annahme.⁶

Gleichung (3) bietet aber noch mehr Informationen: Aus ihr lässt sich das Ausmaß und die Richtung der drohenden Verzerrung ablesen. Je größer (geringer) die unbeobachtete Heterogenität ist, umso geringer (größer) fällt der Betrag der Regressionskoeffizienten aus. Beispielsweise kann ein halb so großer Regressionskoeffizient in Gruppe k gegenüber Gruppe l im Extremfall allein bedeuten, dass die unbeobachtete Heterogenität (σ_k) in der Gruppe k doppelt so hoch ausfällt wie in Gruppe l (σ_l). Nochmals anders ausgedrückt besteht das Problem darin, dass sich keinesfalls der „wahre“ Einfluss α_{ig} der unabhängigen Variablen schätzen lässt, sondern stets nur dieser Einfluss normiert mit der unbekannteten Heterogenität σ_g . Allein das Vorzeichen der Koeffizienten ist zuverlässig interpretierbar. Die Regressionskoeffizienten sind jedoch inhärent mit ihrer Fehlervarianz skaliert. Um dies deutlicher zum Ausdruck zu bringen, wird in Publikationen in der Ökonomie inzwischen oftmals von σ_g als einem „Skalierungsparameter“ gesprochen (z. B. Louviere et al. 2000; Louviere 2001).⁷

Ebenso ist direkt ersichtlich, dass die üblicherweise berichteten Chancenverhältnisse bzw. OR in ähnlicher Weise von diesem Problem belastet sind – schließlich stellen sie eine monotone Funktion der Koeffizienten β dar. Die Erhöhung einer Kovariate x_{ig} um eine Einheit impliziert eine Zunahme der Chancen (*Odds*) um den Faktor $\exp(\beta_{ig})$ (wobei $\exp(\cdot)$ für die Exponentialfunktion steht). Bei kategorialen Variablen messen diese Faktoren zugleich das Chancenverhältnis, das zwischen der jeweiligen Gruppe und Referenzgruppe im Hinblick auf das

interessierende Ereignis besteht (also etwa das zwischen sozialen Klassen bestehende Chancenverhältnis in Bezug auf höhere Bildung).⁸ Genau diese exponierten Koeffizienten sind es, die üblicherweise in den Tabellen als OR berichtet werden. Auch für die OR gilt also: Ihre betragsmäßige Abweichung von dem Zustand, in dem die entsprechende Variable keinen Einfluss ausübt bzw. Chancengleichheit vorliegt (bei OR also vom Wert 1) ist umso größer (geringer), je geringer (größer) die unbeobachtete Heterogenität ist.⁹ Von Allison (1999) wurden diese Aspekte bereits für Karrierechancen von Frauen und Männern (*promotions to associate professors*) dargelegt; und ähnlich bezieht sich eine aktuelle, allein dieser Problematik gewidmete Abhandlung von Carina Mood (2010) im *European Sociological Review* auf das Anwendungsbeispiel von Bildungschancen. Umso mehr erstaunt, dass insbesondere in der Bildungssoziologie zentrale Schlussfolgerungen zu Gruppenunterschieden weiterhin unreflektiert auf Vergleiche von Logit-Koeffizienten gestützt werden.

3. Gruppenvergleiche bei logistischen Regressionen

Was gibt es für Abhilfen – wie lassen sich dennoch valide Gruppenvergleiche durchführen? Eine leicht eingängige, und in den meisten Fällen auch adäquate Möglichkeit besteht darin, statt der absoluten nur relative Koeffizientenwerte miteinander zu ver-

⁸ Die *Odds* geben eine Chance, also Wahrscheinlichkeit relativ zur Gegenwahrscheinlichkeit an, die OR dagegen das relative Chancenverhältnis des interessierenden Ereignisses bei der interessierenden Kategorie (z. B. Mittelklasse m) im Vergleich zur Referenzgruppe (z. B. Dienstklasse d). Formal:

$$OR = \frac{Odds_m}{Odds_d} = \frac{P_m / (1 - P_m)}{P_d / (1 - P_d)}$$

⁹ Zudem soll an dieser Stelle kurz darauf verwiesen werden, dass größere Koeffizientenwerte nach der Aufnahme von zusätzlichen erklärenden Variablen bei binären Regressionen (wiederum anders als bei OLS-Regressionen) nicht zwangsläufig bedeuten müssen, dass der Einfluss der entsprechenden Variablen in der ursprünglichen Schätzung durch diese zusätzlichen Faktoren verdeckt wurde (also eine so genannte *suppressed correlation* vorlag). Der wahrscheinlichere Fall dürfte sein, dass die zusätzlich aufgenommenen Kovariaten die unbeobachtete Heterogenität reduzieren. Dies lässt die Koeffizienten der ursprünglich bereits enthaltenen Variablen zwangsläufig in ihrem Betrag steigen.

⁶ Denn etwa sind theoretisch zahlreiche weitere Einflussfaktoren auf die Bildungsnachfrage anzunehmen, wie etwa die konjunkturelle Lage, Kohortenstärke, Bildungskosten oder die Verfügbarkeit von Stipendien, die ebenfalls zwischen den Kohorten und Regionen variieren dürften. Damit sollte es jeweils in unterschiedlichem Ausmaß mit den berücksichtigten Variablen gelingen, die Bildungsabschlüsse zu erklären. Anders gesagt: Das Ausmaß unbeobachteter Heterogenität unterscheidet sich zwischen den Vergleichsgruppen. Einen groben Hinweis darauf geben Unterschiede in den Pseudo- R^2 -Werten, wie sie bei Hadjar und Berger in den Tabellen für die logistischen Regressionen ausgewiesen sind.

⁷ Auch in der Ökonomie wurde dieser Aspekt lange Zeit übersehen. Das Bekanntwerden hat zu manch Neu-Interpretationen bereits als gesichert geltender Erkenntnisse geführt und Autoren gar zum Ausrufen einer „*rescaling revolution*“ motiviert (Louviere 2001). Diese Revolution hat allerdings bislang kaum die methodischen Grundpfeiler in der Soziologie erschüttert.

gleichen. Hat Merkmal x_d (Zugehörigkeit zur Dienstklasse) in der Gruppe k (Kohorte k) den gleichen relativen Einfluss gegenüber Merkmal x_m (Zugehörigkeit zur Mittelklasse) wie in Gruppe l (Kohorte l)? Ist also etwa bei einem Vergleich der Bildungschancen zwischen verschiedenen Kohorten der Einfluss der Zugehörigkeit zur Dienstklasse stets um den gleichen Faktor größer als derjenige der Zugehörigkeit zur Mittelklasse? Falls nicht, deutet dies auf kohortenspezifische Effektstärken hin. Der einfache *Trick* bei dieser Vorgehensweise besteht darin, durch die Verhältnissetzung der Koeffizienten jeweils den Skalierungsparameter σ_g herauszukürzen (in nachfolgender Gleichung ausführlich auf der linken Seite für die Kohorte k dargestellt, wobei die Indizes d bzw. m die Koeffizienten für die Dienst- bzw. Mittelklasse kennzeichnen, und die Indizes k und l die zwei zu vergleichenden Kohorten):

$$\frac{\beta_{dk}}{\beta_{mk}} = \frac{\alpha_{dk}/\sigma_k}{\alpha_{mk}/\sigma_k} = \frac{\alpha_{dk}}{\alpha_{mk}} \stackrel{?}{=} \frac{\alpha_{dl}}{\alpha_{ml}} \quad (4)$$

Hierauf aufbauend existieren einfache grafische Verfahren, um systematische Gruppenunterschiede aufzudecken (Louviere et al. 2000: Kap. 8). Diese bieten zwar keine Hinweise auf die Signifikanz möglicher Differenzen und sind in seltenen Einzelfällen auch nicht vor Verzerrungen gefeit. Dennoch liefern sie zumindest schon einmal einen ersten Eindruck davon, ob Merkmale in ihrer relativen Effektstärke in einzelnen Gruppen „aus der Reihe tanzen“. ¹⁰ Ferner werden inzwischen verschiedene statistische Testverfahren und Korrekturen für un-

beobachtete Heterogenität empfohlen, die auch bereits in gängigen Statistikprogrammen (wie Stata, SPSS und Limdep) implementiert sind (Mood 2010). Aktuell wurde zudem von Richard Williams (2009) die Verwendung eines *heterogenous choice*-Modells vorgeschlagen, das explizit die Modellierung gruppenspezifischer Heterogenität zulässt.

Interessieren (wie bei Hadjar und Berger) allein deskriptive Vergleiche von Gruppenmittelwerten, gibt es zudem noch einen anderen einfachen Ausweg mittels der Schätzung durchschnittlicher (also über die Beobachtungen gemittelter) Werte. In statistischer Terminologie: Statt der individuellen Werte y_i^* werden deren Erwartungswerte $E(y_i^*)$ geschätzt. Unter den üblichen Gauß-Markov-Annahmen für Regressionen (mit den erklärenden Variablen unkorrelierte Fehlerterme, Erwartungswert der Fehler von 0), reduziert sich damit Gleichung (1) zu:

$$E(y_{ig}^*) = Z_i = \alpha_{0g} + \alpha_{1g}x_{1g} + \dots + \alpha_{pg}x_{pg} \quad (5)$$

Oder vereinfacht in Worten ausgedrückt: Über alle Beobachtungen gemittelt fällt die unbeobachtete Heterogenität nicht ins Gewicht. ¹¹ Eine hiermit korrespondierende Maßzahl für den Einfluss von erklärenden Variablen sind durchschnittliche Marginaleffekte (*averaged marginal effects*, kurz AME): Sie geben an, um wie viele Prozentpunkte sich die Wahrscheinlichkeit des interessierenden Ereignisses *im Mittel* aller (*gruppenspezifischen*) *Beobachtungen* verändert, wenn sich die betreffende erklärende Variable um eine Einheit (*marginal*) erhöht. Damit ist u.E. zugleich ein anschaulicheres Maß für die Effektstärke von Variablen gegeben, als dies mit den eher sperrigen OR der Fall ist. ¹² Formal errechnet sich der AME_j für eine Dummy-Variable x_j folgendermaßen (vgl. Bartus 2005):

$$AME_j = \frac{1}{n} \sum_{i=1}^n \{F(Z_i|x_{ij}=1) - F(Z_i|x_{ij}=0)\} \quad (6)$$

¹⁰ Um zu prüfen, ob die Koeffizienten in der einen Gruppe durchgehend die gleichen Relationen zueinander aufweisen wie in der anderen, werden am einfachsten sämtliche Koeffizienten aus den Regressionsschätzungen der zwei miteinander zu vergleichenden Gruppen in einem Streudiagramm gegeneinander abgetragen. Je eher die Koeffizienten auf einer gemeinsamen Geraden durch den Nullpunkt liegen, umso weniger unterscheiden sie sich in ihren relativen Effektstärken. Die Steigung der Geraden informiert dann über das Ausmaß unbeobachteter Heterogenität, denn wie durch einfache Umformungen von Gleichung (4) ersichtlich ist, beträgt sie σ_k/σ_l . Bei identischer Heterogenität fällt sie also mit der Winkelhalbierenden zusammen (s. Louviere et al. 2000: Kap. 8 für nähere Erläuterungen und ein Beispiel). Diese Methode bietet allerdings nur erste, vorläufige Aufschlüsse, da sich die – zwar unwahrscheinliche – Möglichkeit nicht ganz auszuschließen lässt, dass alle Variablen in der einen Gruppe stets einen um den gleichen Faktor höheren Einfluss aufweisen als in der anderen Gruppe; vgl. Williams 2009 für weitere Fallstricke solch einfacher Verfahrensweisen.

¹¹ Oftmals wird dieser *linearer Prädiktor* – wie in Gleichung (5) angedeutet – einfacher mit dem Buchstaben Z notiert, diese Notation wird nachfolgend noch verwendet.

¹² Diese geben – wie in Fußnote 8 formalisiert – an, um welchen Faktor sich das Verhältnis von Wahrscheinlichkeit und Gegenwahrscheinlichkeit im Vergleich zu dem analogen Wahrscheinlichkeitsverhältnis in der Referenzkategorie ändert, wenn sich die betreffende erklärende Variable um eine Einheit erhöht. Dies ist wenig anschaulich und wird zudem oftmals fälschlich als Veränderung der Auftretenswahrscheinlichkeit des interessierenden Ereignisses interpretiert; vgl. für eine ähnliche Kritik Mood 2010: 80.

Dabei bezeichnet Z_i den linearen Prädiktor für die Beobachtung i , siehe Gleichung (5); $F(\cdot)$ steht für die Verteilungsfunktion, aufsummiert über alle Beobachtungen. Normiert an der Anzahl der Beobachtungen wird also die Differenz der Werte der Verteilungsfunktion des Erwartungswerts für die Bedingung $x_{ij}=1$ von den Werten der Verteilungsfunktion des Erwartungswerts für die Bedingung $x_{ij}=0$ gebildet.

Für Probit-Modelle lässt sich zeigen, dass diese durchschnittlichen Effekte unverzerrt (*konsistent*) geschätzt werden, sofern die nicht im Modell berücksichtigten erklärenden Variablen (die das Problem unbeobachteter Heterogenität verursachen) jeweils Normalverteilungen aufweisen (Wooldridge 2002: 470–472). Für die in den Sozialwissenschaften gängigeren Logit-Modelle und andere Verteilungsformen ergeben sich etwas kompliziertere Zusammenhänge. Allerdings sind in aller Regel die Abweichungen zwischen Logit- und Probit-Schätzungen äußerst klein, so dass es praktisch kaum jemals einen Unterschied machen dürfte, welches binäre Regressionsmodell geschätzt wird – aber eben nicht, welche Parameter für Gruppenvergleiche herangezogen werden.¹³ So kann Carina Mood auch anhand von Simulationen belegen, dass die gemittelten Maße (wie die hier empfohlenen, durchschnittlichen marginalen Wahrscheinlichkeitseffekte) im Falle von Logit-Modellen nur zu einem vernachlässigbar geringen Ausmaß durch unbeobachtete Heterogenität verzerrt sind. Diese Größen lassen sich demnach auch unbedenklich zwischen verschiedenen Modellen, Gruppen, Stichproben oder Kohorten vergleichen (Mood 2010: 78).¹⁴

Bevor die Umsetzung an einem Beispiel gezeigt wird, soll hier noch kurz auf die Möglichkeit inferenzstatistischer Schlüsse eingegangen werden. Gerade für viele theoretische Fragestellungen dürfte oftmals nicht nur interessieren, ob sich Effekte von Merkmalen zwischen Gruppen oder Kohorten unterscheiden, sondern auch, ob diese Unterschiede groß genug ausfallen, um als statistisch signifikant zu gelten. Diese Fragestellung wird oftmals nur *per Augenschein* anhand von Grafiken oder Tabellen begutachtet, nicht aber durch statistische Testgrö-

ßen abgesichert (so etwa auch im Beispiel von Hadjar & Berger 2010). Bei linearen OLS-Regressionen ist es für derartige Signifikanztests üblich, Schätzungen mit Interaktionstermen durchzuführen. Deren Interpretation ist in binären Regressionen allerdings komplex (vgl. Buis 2010) und sie sind überdies ebenfalls von unbeobachteter Heterogenität belastet (Mood 2010).¹⁵

Für bestimmte Fälle, wie dem hier diskutierten Anwendungsbeispiel, lässt sich gleichwohl ein Signifikanztest konstruieren.¹⁶ Die Bedingung dafür ist, dass die kausalen Prozesse in den interessierenden Gruppen unabhängig voneinander ablaufen, also nicht miteinander kovariieren – was bei Bildungsentscheidungen im Kohortenvergleich eine weitgehend realistische Annahme sein dürfte.¹⁷ Schätzt man dann die Einflussstärken getrennt für diese Gruppen (als etwa den Einfluss der sozialen Schicht getrennt für zwei Kohorten), folgt die quadrierte Differenz der aus diesen beiden Schätzungen resultierenden durchschnittlichen Marginaleffekte, normiert durch die Summe aus ihren Varianzen, einer χ^2 -Verteilung mit einem Freiheitsgrad. Formal für den Unterschied des Einflusses des Merkmals x_j in den Gruppen k und l :

$$\frac{(AME_{jk} - AME_{jl})^2}{\text{Var}[AME_{jk}] + \text{Var}[AME_{jl}]} \sim \chi^2(1) \quad (7)$$

Es werden also für die Gruppen k und l getrennte Schätzungen der durchschnittlichen Marginaleffekte vorgenommen, anschließend lässt sich nach Gleichung (7) der Testwert berechnen, um diesen mit

¹⁵ Im Falle von Interaktionstermen ist dies allerdings weniger auf den unbeobachteten Skalierungsparameter zurückzuführen, sondern vielmehr auf die Möglichkeit unterschiedlicher Randverteilungen in den zu vergleichenden Gruppen (Ai & Norton 2003; Buis im Erscheinen).

¹⁶ An dieser Stelle möchten wir uns nochmals ausdrücklich bei Carina Mood für die freundliche Beratung bedanken. Den wertvollen Hinweis auf die nachfolgend beschriebene Vorgehensweise verdanken wir Winfried Pohlmeier.

¹⁷ Problematisch wäre diese Annahme allerdings bei Ausstrahlungseffekten der Verhaltensweisen in einer Kohorte auf die Entscheidungen in anderen Kohorten, etwa wenn besonders große Bildungskohorten aufgrund ihrer starken Besetzung von Arbeitsplätzen nachfolgende Kohorten von höheren Bildungswegen abschrecken – ein Phänomen, das etwas salopp auch unter dem Begriff von *Schweinezyklen* bekannt ist. Ob eine solche Kovariation vorliegt, darauf ergeben sich Hinweise durch die Schätzung der entsprechenden Kovarianzmatrix (gemittelt über alle Beobachtungen und Subgruppen). Sie ist in gängigen Statistikprogrammen einfach anzufordern, etwa in Stata über den Befehl *matrix list e(V)*.

¹³ In den weiter unten demonstrierten Berechnungen von Marginaleffekten bewegen sich die Abweichungen zwischen AME aus Probit- und Logit-Modellen fast ausschließlich auf der dritten Nachkommastelle.

¹⁴ Vorausgesetzt ist allerdings – wie bei jeder Regressions-schätzung –, dass kein *omitted variable bias* vorliegt, also die unbeobachteten Einflussfaktoren nicht mit den im Modell enthaltenen Kovariaten korreliert sind.

dem kritischen Wert für das gewählte Signifikanzniveau abzugleichen („per Hand“ durch Abgleich mit einer tabellierten χ^2 -Verteilung, alternativ lassen sich Signifikanzniveaus durch Statistikprogramme berechnen). Wir werden diese Testmöglichkeit im folgenden Abschnitt demonstrieren.

Für andere Fälle, in denen die Annahme einer fehlenden Kovarianz nicht plausibel ist, können wir an dieser Stelle leider keine einfachen Lösungen anbieten, sondern lediglich auf die Spezialliteratur verweisen – eine verständliche Darstellung der dann einschlägigen Testmöglichkeiten bleibt vorerst ein Desiderat.

4. Praktische Anwendung und Schlussfolgerungen

Wenden wir uns nun dem Anwendungsbeispiel und den Analysen von Hadjar und Berger (2010) zu. Die beiden Autoren untersuchen Bildungsungleichheiten in Westdeutschland, Ostdeutschland und der Schweiz im Kohortenvergleich und beanspruchen eine Antwort auf die Frage zu geben, inwieweit die Bildungsexpansion in kohorten- und länderspezifischer Weise zum Abbau herkunfts- und geschlechtsspezifischer Ungleichheiten geführt hat. Ist die Art der Berechnung der Kohortenunterschiede wirklich von zentraler Bedeutung? Treten die Differenzen zwischen Kohorten und Länderkontexten auch deshalb auf, weil ein jeweils unterschiedliches Ausmaß an unbeobachteter Heterogenität vorliegt – oder, anders gesprochen, die OR mit jeweils unterschiedlichen Skalierungen gemessen wurden?

Vor dem Hintergrund der Analyseziele von Hadjar und Berger, Bildungschancen über Kohorten und Gruppen wie Frauen und Männer zu vergleichen, würde es sich aus statistischer und theoretischer Sicht empfehlen, Quellen unterschiedlicher unbeobachteter Heterogenität, wie die Verwendung unterschiedlicher Datensätze für Deutschland und die Schweiz (oder innerhalb von Deutschland durch Beobachtungen mit unterschiedlichen Panellaufzeiten)¹⁸ soweit wie möglich zu umgehen (wenigstens zur Kontrolle würden sich daher auch Berechnungen mit dem *European Social Survey* oder ALLBUS

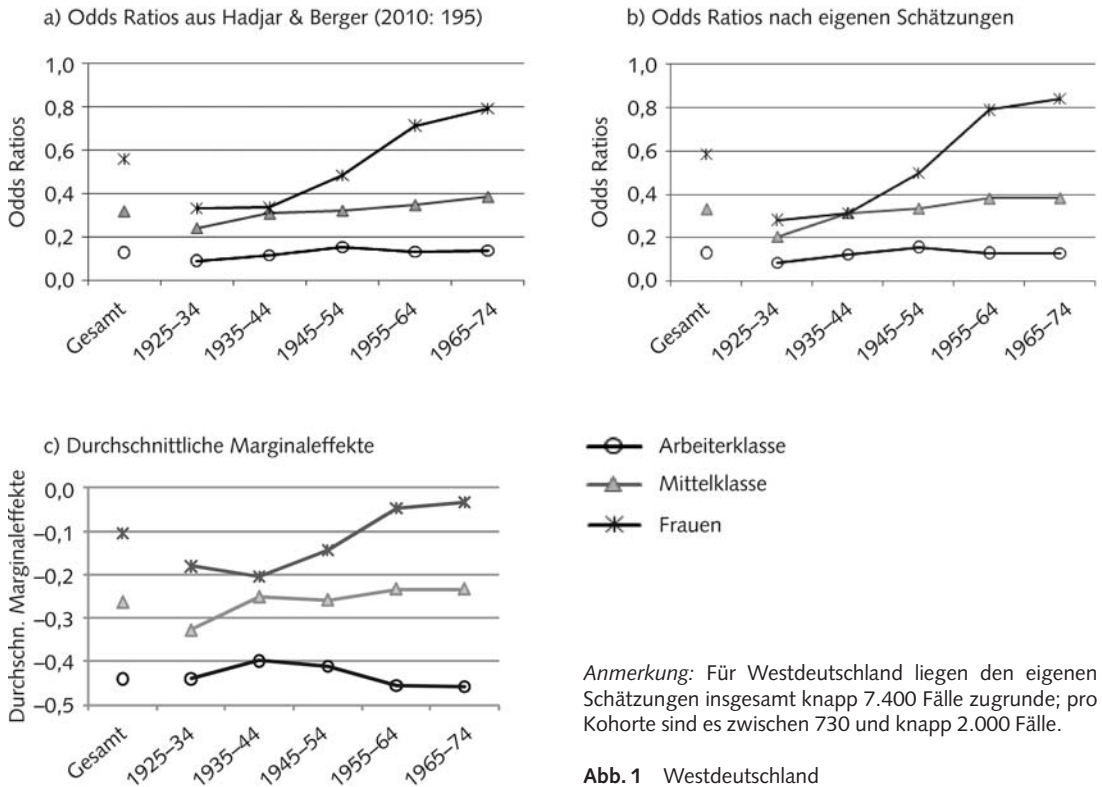
empfehlen). Zudem ist man in derartigen Fällen gut beraten, möglichst viele erklärende Variablen in die Modelle aufzunehmen, um das Ausmaß unbeobachteter Heterogenität gering zu halten.

Um fragliche Verzerrungen exemplarisch zu prüfen, wiederholen wir im Folgenden die anhand des Sozio-ökonomischen Panels (SOEP) vorgenommenen Schätzungen für West- und Ostdeutschland und replizieren dabei so gut wie möglich die Stichprobeneingrenzungen sowie Operationalisierungen aller Variablen. Die Imputationen fehlender Werte können wir nicht nachvollziehen, da hierzu exaktere Angaben zur Verfahrensweise erforderlich wären.¹⁹ Allerdings scheinen die Imputationen für die hier bezweckte Demonstration ohnehin nachrangig. Dennoch wollen wir zunächst kurz auf die Unterschiede eingehen, die sich durch die veränderte Datengrundlage ohne Imputationen ergibt. In der Abbildung 1a (links oben) sind die von Hadjar und Berger für das Geschlecht (oberste Linie) und die soziale Herkunft (mittlere Linie = Mittelklasse; untere Linie = Arbeiterklasse) für Westdeutschland berichteten OR abgetragen (vgl. Hadjar & Berger 2010: 195, Tabelle 4; am linken Rand der Abbildung 1a finden sich die OR für den gesamten Zeitraum; die Linien daneben zeigen den Verlauf über die Kohortenabfolge). Je mehr sich die OR dem Wert 1 bzw. oberen Rand der Grafik annähern, umso stärker gleichen sich die Bildungschancen der jeweiligen sozialen Klasse (bzw. von Frauen) denjenigen der Dienstklasse (bzw. von Männern) an, die hier die Referenzgruppen bilden. Die Abbildung 1b (rechts daneben) zeigt die OR analog geschätzt auf Basis unserer Datengrundlage, die sich wie erwähnt allein durch den Verzicht auf Imputationen unterscheidet. Das Bild ist sehr ähnlich, lediglich der Anstieg der Bildungschancen in der Mittelklasse verläuft nicht ganz so geradlinig wie in den Berechnungen mit Imputationen. Alles in allem gelingt es aber, die Verlaufsmuster sehr gut zu replizieren.

Interessanter ist nun, wie sich die Ergebnisse bei einer alternativen Verwendung von durchschnittlichen Marginaleffekten (AME) darstellen. Diese Maße sind – wie erläutert – bei Probit-Modellen mit normalverteilten Drittvariablen gänzlich und bei Logit-Modellen weitestgehend immun gegenüber Verzerrungen durch unbeobachtete Heterogenität. Wir führen die Schätzungen zunächst anhand von Logit-

¹⁸ Während für Westdeutschland im SOEP Beobachtungen seit dem Jahr 1984 vorliegen, wurde die erste Panelwelle in Ostdeutschland im Jahr 1990 erhoben. Aufgrund dieser unterschiedlich langen Panellaufzeiten dürfte sich das Ausmaß der Panelmortalität, damit aber auch der Selektivität der Stichproben, stark zwischen den Samples für Ost- und Westdeutschland unterscheiden

¹⁹ Die Beschreibungen auf S. 193 lassen darauf schließen, dass allein die soziale Herkunft imputiert wurde – dann irritiert allerdings, dass diese Imputation mitunter auf der Modellvariable „soziale Herkunftsschicht“ basieren soll. Angaben zum verwendeten Schätzmodell fehlen.



Anmerkung: Für Westdeutschland liegen den eigenen Schätzungen insgesamt knapp 7.400 Fälle zugrunde; pro Kohorte sind es zwischen 730 und knapp 2.000 Fälle.

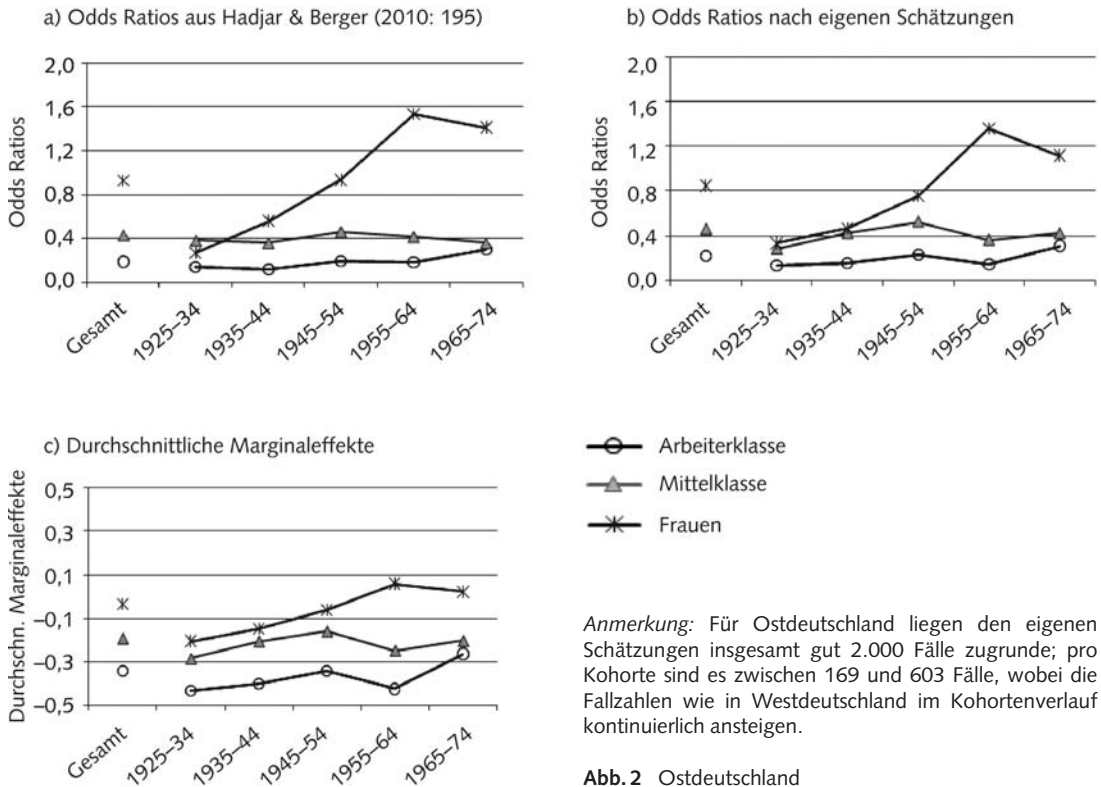
Abb. 1 Westdeutschland

Modellen durch, da diese eine maximale Vergleichbarkeit zu den Analysen von Hadjar und Berger bieten und zudem in den Sozialwissenschaften verbreiteter sind als Probit-Modelle. Die Ergebnisse sind aus der Abbildung 1c (links unten) abzulesen. Die dortige Darstellung basiert auf derselben Datengrundlage wie Abbildung 1b, unterscheidet sich aber durch die Verwendung von AME (statt OR).²⁰ Diese AME geben an, um wie viele Prozentpunkte sich die durchschnittliche Wahrscheinlichkeit eines höheren Bildungsabschlusses in der dargestellten Gruppe von der Wahrscheinlichkeit in der jeweiligen Referenzgruppe unterscheidet. Der am linken Rand abgetragene Gesamteffekt für die Arbeiterklasse von -0,44 bedeutet also beispielsweise, dass Befragte aus der Arbeiterklasse über alle Kohorten

hinweg mit einer um durchschnittlich 44 Prozentpunkte geringeren Wahrscheinlichkeit einen hohen Bildungsabschluss aufweisen als Befragte aus der Dienstklasse. Auch für diese Darstellung gilt dabei: Je stärker sich die Werte dem oberen Grafikrand (hier einem AME von 0) annähern, umso ausgeglichener sind die Bildungschancen. Damit lässt sich in einfacher Weise ablesen, ob die prinzipiellen Verlaufsmuster, also Zu- und Abnahmen der Bildungsungleichheiten im Kohortenverlauf, von dem verwendeten Maß (AME oder OR) abhängen.

In der Tat ist dies der Fall. Während sich die Muster für die Bildungschancen der Frauen in Abbildung 1b und 1c noch am ehesten decken, erweisen sich die Verläufe der Mittel- und Arbeiterklasse in der Darstellung mit durchschnittlichen AME als sehr viel stärker schwankend als bei der Darstellung mit OR. Lediglich im Vergleich der letzten beiden Kohorten ist mit den AME für die Mittelklasse vielleicht noch eine sehr schwache Tendenz zu verbesserten Bildungschancen erkennbar (vgl. Abbildung 1c) – für die Ablesung eines stabilen Trends sind diese wenigen Beobachtungspunkte aber sicher zu wenig. Der von Hadjar und Berger berichtete Bil-

²⁰ Diese wurden mit dem *ado margeff* der Statistiksoftware Stata 11 geschätzt. Dieses Verfahren gewährleistet im Gegensatz zu alternativen Modulen die korrekte Berechnung von marginalen Wahrscheinlichkeitseffekten auch für solche Dummy-Variablen, die mehrere Kategorien einer gemeinsamen zugrunde liegenden Variablen repräsentieren (wie dies hier bei den Klassenkategorien der Fall ist; für Details: Bartus 2005).



Anmerkung: Für Ostdeutschland liegen den eigenen Schätzungen insgesamt gut 2.000 Fälle zugrunde; pro Kohorte sind es zwischen 169 und 603 Fälle, wobei die Fallzahlen wie in Westdeutschland im Kohortenverlauf kontinuierlich ansteigen.

Abb. 2 Ostdeutschland

dungsaufstieg dieser Klasse (etwa im Ergebnisteil mit folgender Passage: „(...) konnten die westdeutschen Mittelklasseangehörigen ihre Position im Bildungssystem stärker und annähernd kontinuierlich verbessern“; Hadjar & Berger 2010: 196) ist somit nicht mehr zu folgern. Substantielle Unterschiede finden sich auch für die Arbeiterklasse: Statt auf konstante oder sogar leicht ansteigende Bildungschancen lassen die AME im Verlauf der letzten vier Kohorten in Westdeutschland deutlich auf sinkende Bildungschancen schließen! Die mit den OR angezeigte Zunahme der Bildungschancen der Mittel- bzw. Arbeiterklasse bis hin zur Kohorte der im Zeitraum 1945 bis 1954 Geborenen scheint also zu einem wesentlichen Teil auch darauf zurückzuführen sein, dass im Kohortenverlauf die unbeobachtete Heterogenität zunimmt, womit die OR unweigerlich betragsmäßig weniger von dem die Chancengleichheit indizierenden Wert 1 abweichen – und dies, obgleich sich die Ungleichheiten anscheinend sogar verstärkt haben.

Für Ostdeutschland sind die Unterschiede nicht ganz so drastisch. Gleichwohl ergibt sich auch hier zumindest für die letzte Kohorte ein anderes Bild

mit den AME statt OR: Vor allem der Anstieg der Bildungschancen für die Arbeiterklasse fällt stärker aus, als dies gemessen mit OR der Fall ist. In dieser Kohorte sind die betragsmäßigen Abweichungen der OR aller Variablen vom Wert 1 (der Chancengleichheit) nach oben verzerrt, was auf eine vergleichsweise geringe unbeobachtete Heterogenität in dieser Kohorte schließen lässt.

Kurz zu erwähnen ist, dass analoge Schätzungen mit Probit-Modellen – wie erwartet – zu identischen Schlussfolgerungen führen. Die Abweichungen bei den geschätzten AME belaufen sich auf maximal einen Prozentpunkt (durchschnittlicher Marginaleffekt von 0,01) und sind damit sicherlich vernachlässigbar.²¹

Kommen wir nun abschließend nochmals auf die Möglichkeit zurück, Unterschiede zwischen Kohorten inferenzstatistisch abzusichern. Nach den Interpretationen von Hadjar und Berger konnte die westdeutsche Mittelschicht ihre Bildungschancen

²¹ Die Schätzungen können auf folgender Internetseite eingesehen werden: <http://www.sozioologie.uni-konstanz.de/hinz/forschung/forschungsprojekte/additional-material>.

Tabelle 1 Logistische Regressionen für die Wahrscheinlichkeit höherer Bildung nach sozialen Klassen in Geburtskohorte 2 und 5, nur Westdeutschland

	Kohorte 2 (1935–1944)		Kohorte 5 (1965–1974)	
	Durchschnittl. Marginaleffekt	Standardfehler	Durchschnittl. Marginaleffekt	Standardfehler
Soziale Klasse (Referenz: Dienstklasse)				
– Arbeiterklasse	–0,399***	–0,0263	–0,459***	–0,0225
– Mittelklasse	–0,252***	–0,0315	–0,234***	–0,0299
Geschlecht (Referenz: Männer)				
– Frauen	–0,204***	–0,0228	–0,0342	0,0202
Fallzahlen	1353		1964	

Anmerkung: Eigene Schätzungen auf Basis des SOEP. Signifikanzniveaus: *** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$.

im Zeitverlauf deutlich und fast durchgehend verbessern. Die hier präsentierten Analysen mit AME legen hingegen eine vorsichtiger Interpretation nahe. Dem Augenschein nach ist allein eine Verbesserung zwischen den ersten beiden Kohorten offensichtlich. Ob sich dieser Trend von der zweiten bis hin zur fünften Kohorte signifikant fortsetzt oder der insgesamt über diesen Zeitraum zu beobachtende leichte Anstieg eher als eine Zufallsschwankung zu deuten ist, lässt sich dagegen ohne statistische Testgrößen nicht mehr ermesen. Zur Prüfung schätzen wir die AME getrennt für die Kohorten 2 und 5. Tabelle 1 zeigt die Ergebnisse, zusammen mit den Standardfehlern. Für Kohorte 2 findet sich ein AME für die Mittelklasse von $-0,252$. Dieser bedeutet, dass Angehörige der Mittelklasse im Vergleich zu denen der Dienstklasse in Kohorte 2 mit einer um 25,2 Prozentpunkte geringeren Wahrscheinlichkeit einen höheren Bildungsabschluss erreichen. In der Kohorte 5 sind es 23,4 Prozentpunkte. Somit hat die so gemessene Bildungsungleichheit in der Mittelklasse von Kohorte 2 zu Kohorte 5 um insgesamt 1,8 Prozentpunkte abgenommen.

Ist dieser Unterschied statistisch signifikant? Die Berechnung gemäß Gleichung (6) ergibt einen χ^2 -Wert von $[(−0,252)−(−0,234)]^2/[(0,0315)^2+(0,0299)^2]=0,172$. Dieser entspricht bei einem zweiseitigen Test mit einem Freiheitsgrad einem p -Wert von 0,678 – was bedeutet, dass sich der Effekt der Mittelklassenzugehörigkeit auf die Bildungschance in Kohorte 5 nicht statistisch signifikant von dem in Kohorte 2 unterscheidet, oder vereinfacht gesprochen, sich die Bildungsungleichheit für die Mittelklasse nicht signifikant reduziert hat.²²

²² Der Test basiert – wie oben allgemein erläutert – auf der Annahme, dass die Prozesse in beiden Kohorten unab-

Die Differenzen zwischen AME und OR mögen *insgesamt* gering erscheinen, sie sind aber im Hinblick auf die in dem diskutierten und ebenso in vielen anderen Aufsätzen verfolgte Strategie – einen Rückschluss auf theoretische Mechanismen anhand von möglichst *präzisen* Deskriptionen zu erzielen – sicher nicht vernachlässigbar. Große gesellschaftliche Prozesse wie die Bildungsexpansion verlaufen immer parallel zu einer Vielzahl von anderen Entwicklungen. Zudem gibt es unzählige Merkmale, in denen sich regionale Kontexte unterscheiden können. Angesichts dessen dennoch kausale Prozesse zu identifizieren und so zur Fortentwicklung von (Bildungs-)Theorien beizutragen, ist eine anspruchsvolle Aufgabe, die auf möglichst solide empirische Messungen angewiesen ist, um nicht zufällige Korrelate mit ursächlichen Effekten zu verwechseln oder tatsächliche Einflüsse aufgrund von Messungenauigkeiten zu verkennen (Müller 1998). Die Frage, ob zeitliche Veränderungen von (Bildungs-) Chancen valide Ergebnisse oder lediglich statistische Artefakte darstellen, wurde schon verschiedentlich aufgeworfen (z. B. Handl 1985) und ebenso wird die Diagnosefähigkeit der Soziologie im Hinblick auf die Folgen der Bildungsexpansion seit langem kritisiert (z. B. Müller 1998). Unser Beitrag soll eine präzisere statistische Analyse anregen und damit auch eine Grundlage für zutreffendere Diagnosen schaffen. Es empfiehlt sich, bei der Interpretation von Logit-Koeffizienten vorsichtiger zu verfahren, adäquatere Maße (wie durchschnittliche Marginaleffekte) und flexiblere Modellierungen

hängig voneinander ablaufen. Die Kovarianz der Koeffizienten von Kohorte 2 und Kohorte 5 beläuft sich bei einer über alle Kohorten gemittelten Schätzung auf $-0,009$, ist somit in der Tat gering.

(wie die von Williams 2009 vorgeschlagenen *heterogeneous choice*-Modelle) zu verwenden. Zu beachten ist zudem, dass unbeobachtete Heterogenität ebenso Gruppenvergleiche auf Basis anderer nichtlinearer Regressionsmodelle belasten kann, also etwa regressionsanalytische Untersuchungen von Verweildauern (Fahrmeir et al. 1996) oder Schätzungen mit multinomialen Logit-Modellen.

Literatur

- Ai, C. & E.C. Norton, 2003: Interaction terms in logit and probit models. *Economics letters* 80: 123–129.
- Allison, P.D., 1999: Comparing Logit and Probit Coefficients Across Groups. *Sociological Methods Research* 28: 186–208.
- Arum, R., A. Gamoran & Y. Shavit, 2007: *More Inclusion Than Diversion: Expansion, Differentiation, and Market Structure in Higher Education*. S. 1–38 in: Y. Shavit, R. Arum & A. Gamoran (Hrsg.), *Stratification in Higher Education. A Comparative Study*. Stanford: Stanford University Press.
- Bartus, T., 2005: Estimation of Marginal Effects Using Margeff. *Stata Journal* 5: 309–329.
- Blossfeld, H.-P., 1993: Changes in Educational Opportunities in the Federal Republic of Germany. S. 51–74 in: Y. Shavit & H.-P. Blossfeld (Hrsg.), *Persistent Inequality*. Boulder: Westview.
- Blossfeld, H.-P. & Y. Shavit, 1993: *Dauerhafte Ungleichheiten. Zur Veränderung des Einflusses der sozialen Herkunft auf die Bildungschancen in dreizehn industrialisierten Ländern*. *Zeitschrift für Pädagogik* 39: 25–52.
- Brüderl, J., 2000: Regressionsverfahren in der Bevölkerungswissenschaft. S. 589–642 in: U. Mueller, B. Nauck & A. Diekmann (Hrsg.), *Handbuch der Demographie*. Band 1. Berlin: Springer.
- Buis, M., im Erscheinen: *The Consequences of Unobserved Heterogeneity in a Sequential Logit Model*. Erscheint in: *Research in Social Stratification and Mobility*.
- Buis, M., 2010: Simple Interpretation of Interactions in Non-Linear Models (Stata tip 87). *Stata Journal* 10: 305–308.
- Fahrmeir, L., A. Hamerle & G. Tutz, 1996: Regressionsmodelle zur Analyse von Verweildauern. S. 301–356 in: L. Fahrmeir, A. Hamerle & G. Tutz (Hrsg.), *Multivariate statistische Verfahren*. Berlin & New York: de Gruyter.
- Hadjar, A. & J. Berger, 2010: Dauerhafte Bildungsungleichheiten in Westdeutschland, Ostdeutschland und der Schweiz: Eine Kohortenbetrachtung der Ungleichheitsdimensionen soziale Herkunft und Geschlecht. *Zeitschrift für Soziologie* 39: 182–201.
- Handl, J., 1985: Mehr Chancengleichheit im Bildungssystem. Erfolg der Bildungsreform oder statistisches Artefakt? *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 37: 698–722.
- Henz, U. & I. Maas, 1995: Chancengleichheit durch die Bildungsexpansion? *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 47: 605–633.
- Louviere, J.J., 2001: What if Consumer Experiments Impact Variances as well as Means? Response Variability as a Behavioral Phenomenon. *Journal of Consumer Research* 28: 506–511.
- Louviere, J.J., D.A. Hensher & J.D. Swait, 2000: *Stated Choice Methods. Analysis and Application*. Cambridge: Cambridge University Press.
- Mood, C., 2010: Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It. *European Sociological Review* 26: 67–82.
- Müller, W., 1998: Erwartete und unerwartete Folgen der Bildungsexpansion. S. 81–112 in: J. Friedrichs, R. Lepsius & K.U. Mayer (Hrsg.), *Die Diagnosefähigkeit der Soziologie. Sonderheft 38 der Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Opladen: Westdeutscher Verlag.
- Müller, W. & D. Haun, 1994: Bildungsungleichheit im sozialen Wandel. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 46: 1–41.
- Schimpl-Neimanns, B., 2000: Soziale Herkunft und Bildungsbeteiligung. Empirische Analysen zu herkunftsspezifischen Bildungsungleichheiten zwischen 1950 und 1989. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 52: 636–669.
- Williams, R.W., 2009: Using Heterogeneous Choice Models To Compare Logit and Probit Coefficients Across Groups. *Sociological Methods & Research* 37: 531–559.
- Wooldridge, J.M., 2002: *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press.

Autorenvorstellung

Katrin Auspurg, geb. 1974 in München. Studium der Sozialen Arbeit sowie der Soziologie, Statistik und Sozialpsychologie an der Ludwig-Maximilians-Universität München. Promotion an der Universität Konstanz. Seit 2006 dort wissenschaftliche Mitarbeiterin am Arbeitsbereich für empirische Sozialforschung.

Forschungsschwerpunkte: Methoden der empirischen Sozialforschung, Soziale Ungleichheit, Familien- und Bildungssoziologie.

Wichtigste Publikationen: *Migration Decisions Within Dual-Earner Partnerships: A Test of Bargaining Theory* (mit M. Abraham & T. Hinz), *Journal of Marriage and Family* 2010; *Herausbildung einer akademischen Elite? Zum Einfluss der Größe und Reputation von Universitäten auf Forschungsförderung* (mit T. Hinz & J. Güdler), *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 2008.

Thomas Hinz, geb. 1962 in München. Studium der Soziologie, Sozialpsychologie und Statistik an der Ludwig-Maximilians-Universität München. Promotion und Habilitation ebenfalls dort. Seit 2004 Professur für empirische Sozialforschung an der Universität Konstanz, 2008 Gastprofessuren an der ETH Zürich und an der Yale University. Forschungsschwerpunkte: Methoden der empirischen Sozialforschung, Arbeitsmarktsoziologie, Wirtschaftssoziologie. Wichtigste Publikationen: Arbeitsmarktsoziologie: Probleme, Theorien, empirische Befunde (mit M. Abraham), Wiesbaden 2008; zuletzt in dieser Zeitschrift: Diffusion einer sozialen Bewegung – lokale Austauschnetzwerke in Deutschland (mit S. Wagner), *ZfS* 39, 2010: 60–80.