

## Choice-Experimente und die Messung von Handlungsentscheidungen in der Soziologie

Katrin Auspurg · Ulf Liebe

**Zusammenfassung:** In umfragebasierten Choice-Experimenten wählen Befragte aus verschiedenen (Handlungs-)Alternativen die von ihnen am meisten präferierte aus. Indem Merkmale dieser Alternativen gezielt variiert werden, lässt sich ihr Einfluss auf die abgefragten Auswahlentscheidungen bestimmen und eine weitaus direktere Prüfung kausaler Zusammenhänge erreichen als mit „herkömmlichen“ Befragungsdaten. Im Gegensatz zu den verwandten Faktoriellen Surveys ist die Methodik direkter auf Handlungs- und Entscheidungstheorien zugeschnitten und bereits eine hohe externe Validität belegt. Im vorliegenden Beitrag werden Choice-Experimente in ihren Grundzügen dargestellt (theoretische Fundierung, Ausgestaltung, Datenerhebung und -auswertung) und vergleichend zu Faktoriellen Surveys diskutiert. Ziel ist es, praktische Handreichungen zu geben und zu vermehrten soziologischen Anwendungen zu motivieren.

**Schlüsselwörter:** Choice-Experimente · Befragung · Handlungstheorien · Random Utility Theory

### Choice-experiments and the measurement of behavioral decisions in sociology

**Abstract:** In survey-based choice experiments respondents choose from various alternative options (for actions) the most preferable one. By systematically varying attributes of the options it is possible to determine their influence on the stated choices. This enables a more direct testing of causal relations than it is possible with “usual” survey data. In contrast to the similar design of factorial surveys the method corresponds better with action and decision theories and a high external validity is already shown. In the paper on hand, choice experiments are introduced in

---

© VS Verlag für Sozialwissenschaften 2011

K. Auspurg (✉)  
Fachbereich Geschichte und Soziologie, Universität Konstanz,  
Universitätsstr. 10, 78464 Konstanz, Deutschland  
E-Mail: katrin.auspurg@uni-konstanz.de

U. Liebe (✉)  
Department für Agrarökonomie und Rurale Entwicklung, Georg-August-Universität Göttingen,  
Platz der Göttinger Sieben 5, 37073 Göttingen, Deutschland  
E-Mail: uliebe@uni-goettingen.de

U. Liebe  
Fachgebiet Soziologie Ländlicher Räume, Universität Kassel,  
Steinstraße 19, 37213 Witzenhausen, Deutschland

their main features (theoretical foundation, design, data collection and analysis) and discussed in comparison to factorial surveys. The aim is to deliver practical tips and thereby motivate increased applications in sociology.

**Keywords:** Choice experiments · Survey · Action theories · Random utility theory

## 1 Einleitung

In Choice-Experimenten (CE) wählen Befragte aus verschiedenen (Handlungs-) Alternativen, so genannten *Choice-Sets*, die von ihnen am meisten präferierte aus. Indem in diesen Alternativen Merkmale (*Attribute*) systematisch variiert werden, lässt sich exakt feststellen, welche Bedeutung ihnen für die abgefragten Auswahlentscheidungen zukommt. Ziehen die Variationen der Merkmale entsprechende Variationen der Entscheidungen nach sich?

Damit versuchen CE ähnlich wie das in der Soziologie weitaus bekanntere Verfahren Faktorieller Surveys (FS) die Vorteile von Experimenten und Umfragen miteinander zu verbinden. Wie bei jedem experimentellen Verfahren ist es besser möglich, theoretisch vermutete kausale Einflüsse zu prüfen. Im Vergleich zu einfachen Itemabfragen werden komplexere Stimuli, damit realistischere und plastischere Urteilsaufgaben geboten und im Gegensatz zur klassischen Laborforschung können Hypothesen vielschichtiger auf der Grundlage größerer (Zufalls-)Stichproben in Bevölkerungsumfragen geprüft werden (Beck und Opp 2001; Nisic und Auspurg 2009). Zudem lässt sich der Anwendungsbereich auf bislang nicht existierende, hypothetische (Handlungs-)Alternativen ausweiten, was für die Evaluation geplanter (politischer) Maßnahmen hilfreich ist. Diese Vorteile haben sich in FS bereits bewährt (Wallander 2009). CE sind im Unterschied zu FS aber weitaus direkter auf die Erhebung von *Entscheidungen* zugeschnitten und stärker an theoretische und statistische Entscheidungsmodelle angebunden. In anderen Disziplinen haben sich CE daher, anders als die dort weitgehend unbekannteren FS, als ein Standardverfahren zur Erhebung von (hypothetischen) Entscheidungen etabliert. Insbesondere für die Ermittlung von Präferenzen und Zahlungsbereitschaften für Güter oder Handlungsoptionen, die (noch) *nicht* auf Märkten gehandelt werden, sind sie einschlägig. Klassische Einsatzgebiete liegen in der Transportwirtschaft (z. B. Hensher 1998) und in der Umwelt- und Gesundheitsökonomik (z. B. Adamowicz et al. 1994; Ryan und Hughes 1997).

Handlungsentscheidungen, die sich nicht auf Marktgüter im engeren Sinne beziehen, stehen aber ebenso in verschiedenen soziologischen Forschungsgebieten im Fokus. Man denke etwa an Bildungsentscheidungen, Entscheidungen über Familiengründungen oder Vertrauensgewährungen. Wie im vorliegenden Beitrag gezeigt werden soll, bieten CE einen vielversprechenden Ansatz zur Prüfung soziologischer Handlungs- und Entscheidungstheorien. Es wird zunächst die Methode in ihren Grundzügen, ihrer Motivation und theoretischen Fundierung erläutert und mit FS kontrastiert (Abschn. 2). Anschließend werden die Erhebungsschritte und Auswertungsverfahren knapp erläutert (Abschn. 3). Das abschließende Resümee (Abschn. 4) weist auf Grenzen der Methodik hin.

## 2 Motivation, Grundprinzip und theoretische Einbettung

Angenommen, es interessiert, was Personen zu medizinischen Vorsorgeuntersuchungen motiviert. Sind Personen bereit, höhere Gebühren und Wartezeiten in Kauf zu nehmen, wenn sie dafür sicherere Diagnosen erhalten? Gilt dies für alle Bevölkerungsschichten gleichermaßen?

Zur Beantwortung könnte man die Wichtigkeit der einzelnen Aspekte über Items erfragen. Dabei würden sich dann aber vermutlich *alle* Aspekte als bedeutsam herausstellen. Mit der Alternative eines Rankings ließe sich die Rangfolge feststellen, das *exakte* Gewicht der einzelnen Entscheidungsmotive bliebe jedoch verborgen. Ökonominen und Ökonomen würden versuchen, die Präferenzen anhand von prozessproduzierten Daten (etwa von Krankenkassen) zu bestimmen. Damit verbunden ist das Problem, dass es sich in verschiedenster Hinsicht nicht um einen *perfekten* Markt mit vollständiger Transparenz und unbeschränktem Wettbewerb handelt. So dürften Patienten nur unzulänglich über den genauen Wert einzelner Behandlungsalternativen informiert sein, und Marktpreise sind aufgrund der starken Subventionen imperfekte Signale für den Nutzen von Alternativen. Zudem kann für neu zu planende Produkte per se noch kein Marktverhalten beobachtet werden (Telser 2002, Kap. 2).

Hier setzt die alternative Erhebungsidee von CE an. In einer gesundheitsbezogenen Anwendung könnten die *Choice-Sets* wie in Abb. 1 aussehen (vgl. die Studie von Ryan et al. 2008). Die Befragten haben sich hier zwischen dem *Status quo*, der in einem Verzicht auf jeglichen Test besteht, und einer Diagnostik zu entscheiden, die zwei verschiedene Testalternativen umfasst („Test A“ oder „Test B“). Die vorgegebenen Attribute der Wahlmöglichkeiten – hier der unterschiedlichen Tests – sind der Informationsumfang, die Wartezeit bis zur Ergebnisverkündung und die Kosten. Die Ausprägungen (Levels) dieser Attribute werden über die Alternativen hinweg variiert (vgl. Tab. 1, für die *Status-quo*-Option nehmen sie stets den Wert Null an). In den Auswertungen lässt sich dann herausfinden, welches Gewicht den einzelnen Attributen in der Entscheidung zukommt. Konkret ist ermittelbar, unter welchen Bedingungen die Befragten überhaupt zu einem Test bereit sind, inwieweit sie willig sind, für sicherere und umfangreichere Tests längere Wartezeiten in Kauf zu nehmen, oder auch wie viel ihnen eine Verkürzung der Wartezeit in Geldeinheiten Wert ist. Es lassen sich die exakten Abwägungen oder *Trade-offs* bestimmen.<sup>1</sup> Zusätzlich ist feststellbar, ob diese Parameter zwischen Befragtengruppen variieren.

Damit weist die Methodik eine hohe Ähnlichkeit zu FS auf. Auch hier werden den Befragten Beschreibungen vorgelegt, die experimentell in ihren Merkmalen variieren (Beck und Opp 2001). Die zwei wesentlichen Unterschiede sind, dass (1) in FS die zu beurteilenden Alternativen (Vignetten) den Befragten sequenziell und nicht simultan wie in den *Choice-Sets* vorgelegt werden und (2) die Befragten dort die Alternativen jeweils für sich auf Ratingskalen beurteilen. Bei einer Umsetzung unseres Beispiels mit einem FS würden die Befragten also nacheinander für einzelne Testbeschreibungen das Ausmaß ihrer Attraktivität oder die Wahrscheinlichkeit ihrer Wahl auf Ratingskalen beurteilen.

---

<sup>1</sup> Ein *Trade-off* liegt vor, wenn eine Person bereit ist, etwas von einem Attribut aufzugeben, um mehr von einem anderen zu erhalten.

	Test A	Test B	Kein Test
Informationen	Einfach	Umfangreich	Keine Informationen
Wartezeit für Testergebnisse	2 Tage	8 Tage	Keine Ergebnisse
Kosten des Tests	20 €	100 €	0 €
<b>Bitte kreuzen Sie denjenigen Test an, den Sie wählen würden.</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**Abb. 1:** Beispiel eines Choice-Sets aus einer Umfrage zur Pränatalen Diagnostik

**Tab. 1:** Attribute und Ausprägungen

Attribute	Erläuterung	Ausprägungen
Informationen	Umfang des Testes, entweder einfacher Test, oder umfangreicher Test, mit dem auch seltene Krankheiten aufge-deckt werden	<i>Keine</i> , einfach, umfangreich
Wartezeit für Testergebnisse	Wartezeit in Tagen bis die Ergebnisse vorliegen	<i>Keine Ergebnisse</i> , 2, 8, 16, 30
Kosten des Tests (€)		<i>0</i> , 20, 100, 250, 500

Ausprägungen für die Status-quo-Alternative „kein Test“ sind kursiv markiert und gelten ausschließlich für diese Alternative

Eine derartige Bewertung einzelner Situationen dürfte deutlich schlechter mit realen Entscheidungen korrespondieren, in denen üblicherweise Alternativen direkt gegeneinander abgewogen werden. Damit erscheint das Antwortformat von FS zumindest für die Messung von Auswahlentscheidungen grundsätzlich weniger geeignet als das von CE. Für die Erhebung von sequenziellen Entscheidungen bleibt abzuwarten, welche Methode valider misst.<sup>2</sup> Zumindest gibt es aber bereits Belege, dass den Befragten explizite Bewertungen einzelner Alternativen schwerer fallen als indirekte Abfragen über Auswahlentscheidungen oder Rankings (Melles 2001, S. 66; Louviere 2006, S. 184). Zudem ist die enge Verbindung zwischen den zu messenden Entscheidungen und dem Erhebungsformat von CE, anders als bei FS, theoretisch begründet.

<sup>2</sup> Sequenzielle Entscheidungen können etwa bei der Jobsuche auftreten, wenn über die Annahme einzelner Angebote entschieden werden muss, ohne dass bereits konkrete Informationen über weitere Angebote vorliegen.

## 2.1 Theoretische Fundierung

Die Methode der CE leitet sich aus der *Characteristics Theory of Value* (CTV, im Wesentlichen Lancaster 1966) sowie der *Random Utility Theory* ab (RUT, McFadden 1974; Manski 1977; für eine tiefere Diskussion: Louviere et al. 2000). Die Kernidee der CTV ist vermeintlich einfach: Nicht ein Gut stiftet per se Nutzen, sondern seine Attribute. Damit eröffnet sich der Blick auf die einzelnen Eigenschaften von Gütern oder Handlungsalternativen. Die RUT geht davon aus, dass der Nutzen von Alternativen ein latentes Konstrukt ist, das (wenn überhaupt) in den Köpfen von Personen existiert.<sup>3</sup> In Form einer indirekten Messung sei es aber prinzipiell möglich, einen signifikanten Anteil des Nutzens zu erklären. Hinzu kommt stets eine unerklärte Komponente, die durch unbeobachtete zusätzliche Attribute, Messfehler oder interindividuelle Varianz bedingt sein kann (vgl. hierzu und dem Folgenden v. a. Amaya-Amaya et al. 2008, S. 15; Louviere et al. 2000). Formal:

$$U_{an} = V_{an} + \varepsilon_{an}. \quad (1)$$

$U_{an}$  entspricht dem latenten, nicht beobachtbaren Nutzen der Wahloption  $a$  für Person  $n$ .  $V_{an}$  ist die systematische oder beobachtbare Komponente und  $\varepsilon_{an}$  der unerklärte Anteil. Einfache Umformungen führen zusammen mit der Grundannahme, dass Personen die Alternative mit dem höchsten Nutzen wählen, zu folgender Formalisierung der Auswahlwahrscheinlichkeit von Option  $a$  statt  $j$  aus dem Choice-Set  $C_n$ :

$$P(a|C_n) = P[(V_{an} + \varepsilon_{an}) > (V_{jn} + \varepsilon_{jn})] = P[(V_{an} - V_{jn}) > (\varepsilon_{jn} - \varepsilon_{an})]. \quad (2)$$

Die Wahrscheinlichkeit von Person  $n$ , die Option  $a$  aus dem Choice-Set  $C_n$  zu wählen, ist gleich der Wahrscheinlichkeit, dass die Summe aus systematischer und willkürlicher Komponente dieser Option für sie größer ist als die entsprechende Summe für die Option  $j$ . Der erklärte Teil wird nun, gemäß der Idee von Lancaster, als additive Funktion der einzelnen Attribute  $X_{an}$  der Wahloption aufgefasst:

$$U_{an} = V(X_{an}) + \varepsilon_{an} = \beta_a + X_{an}\beta_n + \varepsilon_{an}. \quad (3)$$

Je nach Annahme über die Verteilung von  $(\varepsilon_{jn} - \varepsilon_{an})$  in Gleichung (2) resultieren dann Logit- oder Probit-Modelle, um die Einflussgewichte  $\beta_n$  der einzelnen Attribute zu schätzen. In dieser Spezifizierung gilt: Je höher die Ausprägung einer wünschenswerten Eigenschaft ist, desto größer ist der Nutzen dieser Alternative und desto eher wird diese Alternative gewählt.

Wenngleich die theoretische Herkunft der CE eine andere ist, wird speziell mit Gleichung (3) die Nähe zu soziologischen Handlungs- und Entscheidungstheorien offensichtlich. Auch in diesen wird üblicherweise angenommen, dass sich der Nutzen einer

3 Latentes Konstrukt oder latenter Nutzen bezieht sich im vorliegenden Zusammenhang allein auf die Tatsache, dass der tatsächliche Nutzen den Forschenden verborgen bleibt, also nicht unmittelbar beobachtet werden kann.

Alternative additiv zusammensetzt, sich etwa der Nutzen von Weiterbildung aus der erwarteten Bildungsrendite abzüglich den zu erbringenden Investitionen bestimmt, und dass dieses (latente) Nutzenkalkül die getroffenen Entscheidungen *erklärt*. CE bedingen durch ihren theoretischen Hintergrund eine engere Verzahnung von Theorie und Empirie, als dies bei herkömmlichen Itembatterien oder FS der Fall ist. Denn Anwender von FS beschränken sich in der Regel auf eine theoretische Ableitung der relevanten Urteils- oder Entscheidungsfaktoren (z. B. Kosten und Nutzen), ohne ihre Verbindung zu den Antworten explizit zu formalisieren. Es wird lediglich postuliert, dass Kosten (Nutzen) die Attraktivität von Alternativen reduzieren (erhöhen), ohne aber das Erhebungsformat und Auswertungsmodell näher theoretisch zu begründen.

### 3 Durchführung

Es sollen nochmals die zentralen Bestandteile von CE erinnert werden: Die Befragten erhalten *Choice-Sets*, also tabellarisch gegenüber gestellte Alternativen, aus denen sie die meist präferierte auswählen sollen. Die Alternativen setzen sich aus verschiedenen Attributen zusammen, die in ihren Ausprägungen oder Levels variieren. Gängig ist es überdies, den einzelnen Befragten mehrere Choice-Sets zu präsentieren und dazu mit unterschiedlichen Fragebogenversionen oder *Blöcken* zu arbeiten.

#### 3.1 Spezifikation der Choice-Sets

Zunächst ist zu überlegen, welche und wie viele Alternativen den Befragten in einem gemeinsamen Choice-Set vorgelegt werden (Amaya-Amaya et al. 2008; Liebe 2007; Louviere et al. 2000 zum Folgenden). Die Alternativen können mit inhaltlichen (z. B. „einfacher Test“, „umfangreicher Test“, sogenannte *labeled alternatives*) oder allgemeinen Labels überschrieben werden (z. B. „Test A“, „Test B“, sogenannte *generic alternatives*). In beiden Varianten ist es möglich, eine Alternative mit dem Status-quo zu besetzen, der über die Choice-Sets hinweg konstant bleibt und die derzeitige Situation beschreibt („kein Test“). Die Berücksichtigung einer Status-quo- oder Opt-out-Alternative wird aufgrund der stärkeren Realitätsnähe empfohlen und um keine ungewollten Entscheidungen zu erzwingen. Was die Anzahl der Alternativen betrifft, liegen gute Erfahrungen mit bis zu vier Optionen vor. Unser Beispiel in Abb. 1 beinhaltet drei generische Alternativen, wobei die dritte Alternative, „kein Test“, den Status-quo darstellt.

Als nächstes sind die Attribute festzulegen. Dies geht Hand in Hand mit einer Spezifikation des unterstellten Nutzenmodells und erfolgt damit in sozialwissenschaftlichen Anwendungen theoriegeleitet. Als empirisch bewährte Orientierungsgröße gelten fünf bis maximal neun Attribute. Unser Beispiel in Abb. 1 umfasst drei Attribute („Informationen“, „Wartezeit“ und „Kosten“).

Nachdem die Attribute bestimmt wurden, steht die Festlegung ihrer Levels an. Für die Modellierung nicht-linearer Zusammenhänge sind mindestens drei Levels erforderlich. Mehr sollten es nur im Ausnahmefall sein (wie im Beispiel beim Preisattribut). Zudem sollte auf eine zahlenmäßige Ausgewogenheit der Levels geachtet werden, um nicht einzelnen Attributen überproportionales Gewicht zu verleihen (Wittink et al. 1982). Ein-

setzbar sind sowohl quantitative Levels (z. B. Wartezeit in Tagen) als auch qualitative („kurze“, „lange“ Wartezeit). Statistisch vorzuziehen sind erstere, denn sie implizieren weniger zu schätzende Parameter. In jedem Fall sollten die Attribute genau beschrieben werden und die Ausprägungen realistisch sein. Zur Beschreibung können auch einleitende Informationen oder eine grafische Darstellung durch Symbole und Bilder verwendet werden.

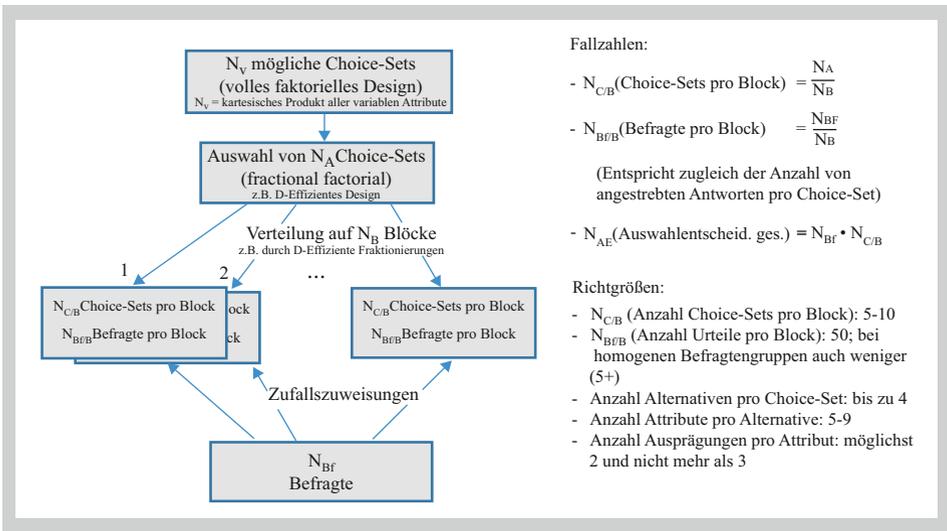
### 3.2 Festlegung des experimentellen Designs

Anschließend ist die konkrete Zusammenstellung der Choice-Sets und Alternativen vorzunehmen. Das Universum aller möglichen Kombinationen, das so genannte *volle faktorielle Design (full factorial)*, bildet sich durch Kreuzung aller Levels und Alternativen. In unserem Beispiel ergeben sich bei zwei Attributen à vier Levels und einem Attribut à zwei Levels  $4 \times 4 \times 2 = 32$  mögliche Kombinationen für die Bildung einzelner Alternativen. Zudem besteht jedes Choice-Set aus zwei variablen Alternativen. Damit ergeben sich insgesamt  $(4 \times 4 \times 2)_{\text{TestA}} \times (4 \times 4 \times 2)_{\text{TestB}} = 1024$  mögliche Choice-Sets (wobei hier inhaltlich unsinnige Kombinationen noch zugelassen sind). Der Umfang des *full factorial* steigt exponentiell mit der Anzahl an Levels an, was die eben ausgesprochene Empfehlung von sparsamen Anzahlen erklärt. In der Regel ist es dennoch nicht praktikabel, den Befragten alle Choice-Sets vorzulegen. Es wird daher mit einer Auswahl, einem *fractional factorial* gearbeitet.

Durch Auswahlen kommt es unvermeidlich zu Korrelationen oder Konfundierungen von Attributen. Um damit verbundene Qualitätseinbußen gering zu halten (Korrelationen verringern die Genauigkeit, mit der sich separate Einflüsse der einzelnen Attribute schätzen lassen), sollte unbedingt eine bewusste Selektion vorgenommen werden (Steiner und Atzmüller 2006). Als besonders vorteilhaft gelten fraktionalisierte Designs, welche folgende vier Aspekte optimieren (Huber und Zwerina 1996): (1) die maximale Unkorreliertheit der Attribute (*orthogonality*), (2) die zahlenmäßige Ausgewogenheit ihrer Ausprägungen (*level balance*), (3) die minimale Überschneidung der Ausprägungen in einem gemeinsamen Choice-Set (*minimal overlap*), und (4) möglichst gleich aufliegende Nutzenwerte der Alternativen eines Choice-Sets (*utility balance*). Alle vier Kriterien erhöhen den Informationsgehalt, indem sie die unabhängige Varianz der Attribute maximieren und die Befragten zu stärkeren Abwägungen zwingen.

Für die Zusammenstellung solcher *fractional factorials* gibt es vorgefertigte Experimentalpläne (Street und Burgess 2007). Alternativ einsetzbare Computeralgorithmen eröffnen höhere Gestaltungsspielräume, weil sich etwa unlogische Kombinationen und interessierende Wechselwirkungen von Attributen gezielt berücksichtigen lassen. Sie sind mit hoher Güte mit den als „Free-Ware“ beziehbaren Zusatzmakros von Warren Kuhfeld zur Statistik-Software SAS umsetzbar (Kuhfeld 2009).<sup>4</sup> Ideale Stichprobenumfänge las-

4 Es empfiehlt sich, dort auf so genannte D-effiziente oder -optimale Fraktionen zurückzugreifen und die Attribute der einzelnen variablen Alternativen eines Choice-Sets (also hier von „Test A“ und „Test B“) bei der Designbildung wie separate Attribute zu behandeln. Im englischsprachigen Raum liegt mit *Ngene* mittlerweile eine Spezial-Software für Choice-Experimente vor, mit der sich einfache und sehr komplexe Designs umsetzen lassen (<http://www.choice-metrics>).



Anmerkung: Die Berechnung der Fallzahlen ist an eine entsprechende Darstellung für Faktorielle Surveys bei Beck und Opp (2001) angelehnt.

**Abb. 2:** Ablaufschritte und Fallzahlen

sen sich ebenfalls mit diesen Algorithmen ermitteln. Etwa 100 Choice-Sets bilden eine erste Richtgröße.

Als nächstes sind die ausgewählten Choice-Sets zu Blöcken zu gruppieren, welche dann einzelnen Befragten vorgelegt werden. Der Hauptgrund für die Verwendung von mehreren Choice-Sets pro Befragten sind Einsparungen an Befragungsaufwand und Kosten, da bereits mit verhältnismäßig wenigen Befragten hinreichende Fallzahlen für multivariate Auswertungen erzielt werden. Die komplette Auswahl von Choice-Sets ist aber, selbst wenn sie bereits durch ein fraktionalisiertes Design reduziert wurde, in der Regel immer noch zu umfangreich für einzelne Befragte. Die praktische Lösung besteht in der Bildung unterschiedlicher Fragebogenversionen bzw. Blöcke. Wie viele Blöcke angefertigt werden, ist abhängig von der anvisierten Anzahl an Entscheidungen, Befragten und Choice-Sets. Ein oftmals zu findender Richtwert liegt bei 50 Befragten pro Block (Bennett und Adamowicz 2001, S. 59). Speziell bei homogenen Befragtensamples erscheinen aber bereits fünf Befragte aufwärts hinreichend.<sup>5</sup> Zu beachten ist, dass bei durchschnittlich komplexen Alternativen (7±2 Attribute) ab etwa dem zehnten Set mit Lern- und Ermüdungserscheinungen zu rechnen ist (Bradley und Daly 1994, S. 180; Caussade et al. 2005, S. 631 f.). Wurde die Entscheidung über die Größe der Blöcke getroffen, bedarf es einer Technik zu ihrer Zusammenstellung. Hierzu können wiederum die genannten Computer-Algorithmen genutzt werden. Abbildung 2 fasst die bis hierhin durchzufüh-

com). In dieser Software ist der aktuelle Forschungsstand zu experimentellen Designs bereits implementiert, nämlich weitere Effizienzgewinne durch die Berücksichtigung von Vorinformationen zu den Koeffizientenwerten  $\beta_n$  (priors) zu erzielen (Street und Burgess 2007).

5 Es muss nur in jedem Fall sichergestellt werden, dass alle ausgewählten Choice-Sets beurteilt werden und die Blöcke nicht mit Merkmalen der Befragten korrelieren.

den Schritte und Kalkulationen der Fallzahlen nochmals zusammen; es handelt sich um Richtlinien, die für den Einzelfall anzupassen sind (z. B. anhand von Vorstudien).

### 3.3 Datenerhebung

Als Befragungsmodus kommen eine persönliche (*face-to-face*), schriftliche (*paper-and-pencil*) oder Online-Erhebung in Betracht. Bei einer Vorabzusendung der Unterlagen kann zudem telefonisch befragt werden. Programmierungen der Fragebögen haben den Vorteil, dass zufällige Abfolgen der Choice-Sets, die Reihenfolge- oder *Halo*-Effekten vorbeugen (Bradley und Daly 1994), einfach zu implementieren sind. Unerlässlich ist in jedem Falle eine zufällige Verteilung der Blöcke auf Befragte, da nur so der experimentelle Charakter (Unabhängigkeit von Befragtenmerkmalen und experimentellen Stimuli) gesichert ist. Diese Randomisierung kann durch eine entsprechende Programmierung oder zufällige Mischung von Papier-Fragebögen erreicht werden.

Die Aufgabe der Befragten ist es dann, pro Choice-Set die von ihnen am meisten präferierte Alternative auszuwählen. Sinnvoll ist es, zusätzlich soziodemografische Merkmale, themenspezifische Einstellungen etc. zu erfragen.

## 4 Datenauswertung

### 4.1 Datenstruktur

Einen Auszug aus der Datenmatrix für unser Medizin-Beispiel zeigt Tab. 2. Jede gewählte und nicht gewählte Alternative (hier Variable „Alt“) eines Choice-Sets erhält im Datensatz eine Zeile, in der neben der Entscheidung der Befragten (*gewählt* versus *nicht-gewählt*, hier Variable „Wahl“) ebenfalls die einzelnen Ausprägungen der Attribute (hier z. B. die Variable „Zeit“ für Wartezeit) und soziodemografische sowie andere Merkmale der Befragten (hier z. B. Alter) enthalten sind.<sup>6</sup> Es liegt ein Datensatz mit drei Ebenen vor, bei dem die unterste Ebene die Auswahlentscheidungen bilden, die mittlere die Choice-Sets und die oberste die einzelnen Befragten. Haben beispielsweise 300 Befragte jeweils sechs Choice-Sets mit drei Alternativen bewertet, dann ergeben sich  $300 \times 6 \times 3 = 5400$  Datenzeilen.

### 4.2 Auswertungsverfahren

Für die Auswertung von CE können Standard-Verfahren zur Analyse diskreter Entscheidungen genutzt werden (z. B. Maier und Weiss 1990). Handelt es sich um lediglich zwei Alternativen in den Choice-Sets, sind binäre Logit- oder Probit-Modelle gängig, bei höheren Anzahlen an Alternativen konditionale Logit-Modelle (KL, Louviere et al. 2000, S. 13). KL modellieren den Einfluss von Attributen, die zwischen Alternativen variieren,

---

6 Die Attribute brauchen bei der Dateneingabe nicht erfasst zu werden. Es reicht die Erfassung des Blocks und der Nummern der Choice-Sets, um die Attribute aus dem experimentellen Designplan dazu spielen zu können.

auf die Auswahlwahrscheinlichkeit von Alternativen – und dies unabhängig davon, um welche Alternative es sich handelt. In unserem Anwendungsbeispiel würde also beispielsweise ermittelt, welchen Effekt die Wartezeit auf die Auswahlwahrscheinlichkeit einer Behandlungsform hat, unabhängig davon ob es sich nun um „Test A“, „Test B“ oder „keinen Test“ handelt. KL-Modelle bieten eine hohe Übersichtlichkeit der Ergebnisse, bezahlen diese aber mit den oftmals unrealistischen Annahmen, dass die Einflüsse unabhängig von den konkreten Alternativen sind und eine Unabhängigkeit von irrelevanten Alternativen besteht (*Independence of Irrelevant Alternatives, IIA*; McFadden 1974, S. 109; Long und Freese 2006) – also das Verhältnis der Auswahlwahrscheinlichkeit von zwei Alternativen unabhängig vom Auftreten anderer Alternativen ist. Dies ist speziell dann unrealistisch, wenn sich einzelne Alternativen in einem Choice-Set grundsätzlich stärker ähneln als die übrigen. In solchen Fällen ist es oftmals schlüssiger, explizit von einer sequenziellen Struktur des Entscheidungsverhaltens auszugehen (etwa zunächst Entscheidung für einen medizinischen Test ja/nein, und erst anschließend Wahl zwischen verschiedenen Testmöglichkeiten). Solche Entscheidungsstrukturen bilden *Nested-Logit*-Modelle ab. Flexibilität zur Umgehung der IIA- und weiterer Annahmen bieten zudem komplexere Logit-Modellierungen, die auch den Panelcharakter der Daten, also mehrere Bewertungen pro Befragten, berücksichtigen (z. B. *mixed-Logit*-Modelle; Hensher et al. 2005; Louviere et al. 2000; detailliert Long und Freese 2006). Letzteres ermöglicht korrekte Schätzungen trotz Verletzung der Unabhängigkeitsannahme (die einzelnen Befragten nehmen mehrere Bewertungen vor, weshalb die einzelnen Urteile nicht unabhängig voneinander sind).

Es empfiehlt sich die Aufnahme sogenannter *alternativenspezifischer Konstanten*. Sie dienen als Platzhalter für systematische Einflüsse der einzelnen Alternativen, die (noch) nicht adäquat modelliert sind. Oftmals ist beispielsweise eine starke Tendenz zum Status-quo zu beobachten, die sich nicht allein über die Attribute erklären lässt. Die Konstanten werden als Dummy-Variablen gebildet, die immer den Wert Eins annehmen, falls die jeweilige Handlungsalternative vorliegt (z. B. Test A), ansonsten den Wert Null. Bei  $j$  Handlungsalternativen können  $j - 1$  Konstanten in das Auswertungsmodell aufgenommen werden. In unserer Beispiel-Datenmatrix in Tab. 2 bilden Alt1, Alt2 und Alt0 solche alternativenspezifischen Konstanten. Falls sich Handlungsalternativen stark ähneln, ist es oft zweckmäßig, ihnen eine identische Konstante zuzuordnen. Eine Besonderheit von Regressionsmodellen für diskrete Entscheidungen ist zudem, dass Befragtenmerkmale nicht direkt als Variablen einbezogen werden können. Sie gehen über Interaktionsterme mit den Attributen oder mit den alternativenspezifischen Konstanten in die Auswertungsmodelle ein. Signifikante Interaktionen mit den alternativenspezifischen Konstanten bedeuten, dass Befragtengruppen grundsätzlich eine verschieden starke Präferenz für einzelne Alternativen aufweisen; signifikante Interaktionen mit Attributen zeigen dagegen an, dass einzelnen Attributen eine unterschiedliche Bedeutung zugemessen wird.<sup>7</sup>

Eine Besonderheit bilden zudem Kennziffern wie die bereits häufiger angesprochenen Trade-offs. Einfache Umformungen des RUT-Modells verdeutlichen (s. Abschn. 2), dass sich *implizite Preise* oder *Teilnutzenwerte* (TNW; *part-worth*) nicht-monetärer Attribute  $k$

<sup>7</sup> Interpretationen von Interaktionstermen in logistischen Regressionen sind allerdings nicht unproblematisch, s. Auspurg und Hinz 2011 und Mood 2010 für mehr Details.

**Tab. 2:** Auszug aus der Datenmatrix

ID	Cset	Alt	Alt1	Alt2	Alt3	Info	Zeit	Kost	Wahl	Alter
2	4	1	1	0	0	1	16	500	1	30
2	4	2	0	1	0	0	2	250	0	30
2	4	3	0	0	1	2	0	0	0	30
2	5	1	1	0	0	1	8	20	0	30
2	5	2	0	1	0	0	30	100	1	30
2	5	3	0	0	1	2	0	0	0	30
3	4	1	1	0	0	1	16	500	0	40
3	4	2	0	1	0	0	2	250	0	40

ID kennzeichnet die einzelnen Befragten, Cset meint Nummer des Choice-Sets; Alt bezieht sich auf die jeweilige Alternative im Choice-Set; Alt1, Alt2, Alt3 sind alternativenspezifische Konstanten mit der Ausprägung 1, wenn die jeweilige Alternative vorliegt; Info ist eine ordinal kodierte Variable für das Informationsattribut mit den Ausprägungen 1=einfach, 2=umfangreich und 0=keine Informationen; Zeit bezieht sich auf das Attribut „Wartezeit für die Testergebnisse“; Kosten auf das Attribut „Kosten des Tests“; Wahl gibt die gewählte Alternative pro Choice-Set an, und Alter bezieht sich auf das Alter der befragten Person

durch Gewichtung ihres Koeffizienten  $\beta_k$  mit dem Koeffizientenwert eines monetären Attributs  $\beta_m$  bestimmen lassen (multipliziert mit -1):

$$\text{TNW} = - \frac{\beta_k}{\beta_m}$$

So wäre in unserem Beispiel berechenbar, wie hoch die mittlere Zahlungsbereitschaft für eine eintägige Verkürzung der Wartezeit ist ( $\beta_k$  würde in diesem Falle den Koeffizienten der in Tagen gemessenen Wartezeit darstellen;  $\beta_m$  den Koeffizienten für die Kosten des Tests). Ähnlich lassen sich weitere anschauliche Kenngrößen ermitteln, wie die für bestimmte Produkte zu erwartenden Marktanteile oder Wohlfahrtsmaße, welche den gesellschaftlichen „Gewinn“ von (Politik-)Maßnahmen beziffern (ausführlich Bennett und Adamowicz 2001).

## 5 Diskussion und Resümee

CE können den Erkenntnisfortschritt immer dann anreichern, wenn Entscheidungen zwischen diskreten Alternativen interessieren. Das Verfahren liefert tiefgehende Informationen zum Gewicht einzelner Entscheidungsfaktoren und eignet sich somit für die gezielte Prüfung von (soziologischen) Handlungs- und Entscheidungstheorien. Im Vergleich zu FS bestechen CE durch eine stringenter theoretische Fundierung und stärkere Korrespondenz zwischen den interessierenden Entscheidungen und Antwortformaten.

Gleichwohl sind auch Nachteile zu diskutieren. Zu diesen zählt die vergleichsweise komplexe Urteilsaufgabe. Bekannte Folgen von kognitiven Überforderungen sind geringere Konsistenzen des Antwortverhaltens, der Rückgriff auf Entscheidungsheuristiken (z. B. Ausblendung einzelner Merkmale), vermehrte Wahlen der Status-quo-Option oder

völlige Antwortverweigerungen (Auspurg et al. 2009; Deshazo und Fermo 2002; Hensher et al. 2005; Meyerhoff und Liebe 2009; Swait und Adamowicz 2001). Derartige Effekte sind jedoch durch Einhaltung der oben genannten methodischen Richtlinien gering zu halten (Hensher 2006). Heuristiken sind zudem nicht notwendig methodische Artefakte; sie können im Gegenteil mit entsprechenden Strategien bei realen Entscheidungen korrespondieren (Gigerenzer und Todd 1999). Gleichwohl sollten Ergebnisse wie Zahlungsbereitschaften sicher nicht „auf den Cent“ genau ausgelegt werden.

Der Hauptkritikpunkt an CE bezieht sich sicher aber darauf, dass lediglich hypothetische und keine realen Entscheidungen erfasst werden. Eine mangelnde externe Validität wird insbesondere Anreize zu sozialer Erwünschtheit und einem strategischen Antwortverhalten zugeschrieben (Liebe 2007, S. 134). Die indirekte Bewertungsmethode der Choice-Sets gilt jedoch zumindest als immun gegenüber solchen Effekten als direkte Abfragen von Präferenzen (Louviere et al. 2000, S. 351 f.). Zudem können CE von einer, gegenüber FS wesentlich ausgereifteren, Methodenforschung profitieren. Die Ergebnisse zur externen Validität weisen insgesamt auf eine erstaunlich hohe Übereinstimmung der hypothetischen mit realen Entscheidungen hin, und dies selbst bei stark sozial erwünschten Verhaltensweisen wie der Bereitschaft, Aufpreise für ökologische Produkte zu zahlen (Blamey und Bennett 2001; Carlsson und Martinsson 2001; Louviere et al. 2000; Louviere und Timmermanns 1992; Telser und Zweifel 2007).

Die recht aufwendige Erstellung der Choice-Sets dürfte sich auf der Erhebungsseite in der Regel durch einen vergleichsweise geringen Aufwand auszahlen. So lassen sich bereits mit wenigen Befragten aussagekräftige Fallzahlen für multivariate Analysen erzielen, und es entfällt eine groß angelegte Suche nach Spezialpopulationen, wenn seltene Entscheidungen, wie etwa Entscheidungen über berufliche Fernumzüge interessieren (hierzu existiert bereits eine Pilotstudie der Autoren). Die Anforderung, ein exaktes Nutzenmodell aufzustellen, diszipliniert den Forscher, seine Hypothesen vorab genau zu spezifizieren. Im Sinne eines effizienten Erkenntnisgewinns und Umsetzung einer analytischen Soziologie (Hedström 2005) ist dies als weiterer Vorteil zu werten. Die mit CE gewonnenen Daten ersetzen zwar nicht reale Beobachtungen, sie bieten aber im Rahmen einer kumulativen Forschungsstrategie eine sinnvolle Ergänzungsmöglichkeit.

**Danksagung:** Die Autoren danken Martin Abraham, Thomas Hinz sowie Jürgen Meyerhoff für wertvolle Rückmeldungen zu einer früheren Version.

## Literatur

- Auspurg, Katrin, und Thomas Hinz. 2011. Gruppenvergleiche bei Regressionen mit binären abhängigen Variablen – Probleme und Fehleinschätzungen am Beispiel von Bildungschancen im Kohortenverlauf. *Zeitschrift für Soziologie* 40:62–73.
- Auspurg, Katrin, Thomas Hinz, und Stefan Liebig. 2009. Komplexität von Vignetten, Lerneffekte und Plausibilität im Faktoriellen Survey. *Methoden, Daten, Analysen* 3:59–96.
- Adamowicz, Wiktor, Jordan J. Louviere, und Michael Williams. 1994. Combining revealed and stated preference methods for valuing environmental amenities. *Journal of Environmental Economics and Management* 26:271–292.

- Amaya-Amaya, Mabel, Karen Gerard, und Mandy Ryan. 2008. Discrete choice experiments in a nutshell. In *Using discrete choice experiments to value health and health care*, Hrsg. Mandy Ryan, Karen Gerard, und Mabel Amaya-Amaya, 13–46. Dodrecht: Springer.
- Beck, Michael, und Karl-Dieter Opp. 2001. Der faktorielle Survey und die Messung von Normen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 53:283–306.
- Bennett, Jeff, und Vic Adamowicz. 2001. Some fundamentals of environmental choice modelling. In *The choice modelling approach to environmental valuation*, Hrsg. Jeff Bennett und Russell Blamey, 37–72. Cheltenham Northampton: Edward Elgar.
- Blamey, Russell, und Jeff Bennett. 2001. Yea-saying and validation of a choice model of green product choice. In *The choice modelling approach to environmental valuation*, Hrsg. Jeff Bennett und Russell Blamey, 179–201. Cheltenham Northampton: Edward Elgar.
- Bradley, Mark, und Andrew Daly. 1994. Use of the logit scaling approach to test for rank-order and fatigue effects in stated preference data. *Transportation* 21:167–184.
- Carlsson, Fredrik, und Peter Martinsson. 2001. Do hypothetical and actual marginal willingness to pay differ in choice experiments? Application to the valuation of the environment. *Journal of Environmental Economics and Management* 41:179–192.
- Caussade, Sebastian, Juan De Dios Ortúzar, Luis I. Rizzi, und David A. Hensher. 2005. Assessing the influence of design dimensions on stated choice experiment estimates. *Transportation Research Part B* 39:621–640.
- Deshazo, J. R., und German Fermo. 2002. Designing choice sets for stated preference methods: The effects of complexity on choice consistency. *Journal of Environmental Economics and Management* 44:123–143.
- Gigerenzer, Gerd, und Peter M. Todd. 1999. *Simple heuristics that make us smart*. New York: Oxford University Press.
- Hedström, Peter. 2005. *Dissecting the social. On the principles of analytical sociology*. Cambridge: University Press.
- Hensher, David A. 1998. Establishing a fare elasticity regime for urban passenger transport. *Journal of Transport Economics and Policy* 32:221–246.
- Hensher, David A. 2006. How do respondents handle stated choice experiments? Attribute processing strategies under varying information load. *Journal of Applied Econometrics* 21:861–878.
- Hensher, David A., John Rose, und William H. Greene. 2005. The implications on willingness to pay for respondents ignoring specific attributes. *Transportation* 32:203–222.
- Huber, Joel, und Klaus Zwerina. 1996. The importance of utility balance in efficient choice designs. *Journal of Marketing Research* 33:307–317.
- Kuhfeld, Warren F. 2009. *Marketing research methods in SAS. Experimental design, choice, conjoint and graphical techniques*. Cary: SAS Institute.
- Lancaster, Kelvin J. 1966. A new approach to consumer theory. *The Journal of Political Economy* 74:132–157.
- Liebe, Ulf. 2007. *Zahlungsbereitschaft für kollektive Umweltgüter. Soziologische und ökonomische Analysen*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Long, Scott J., und Jeremy Freese. 2006. *Regression models for categorical dependent variables using stata*. College Station: Stata Press.
- Louviere, Jordan J. 2006. What you don't know might hurt you: Some unresolved issues in the design of discrete choice experiments. *Environmental & Resource Economics* 34:173–188.
- Louviere, Jordan J., und Harry J. P. Timmermans. 1992. Testing the external validity of hierarchical conjoint analysis models of recreational destination choice. *Leisure Sciences* 14:179–194.
- Louviere, Jordan J., David A. Hensher, und Joffre D. Swait. 2000. *Stated choice methods. Analysis and application*. Cambridge: Cambridge University Press.
- Maier, Gunther, und Peter Weiss. 1990. *Modelle diskreter Entscheidungen. Theorie und Anwendungen in den Sozial- und Wirtschaftswissenschaften*. Wien: Springer.
- Manski, Charles F. 1977. The structure of random utility models. *Theory and decision* 8:229–254.

- McFadden, Daniel. 1974. Conditional logit analysis of qualitative choice behavior. In *Frontiers in econometrics*, Hrsg. Paul Zarembka, 105–142. New York: Academic.
- Melles, Torsten. 2001. *Framing-Effekte in der Conjoint-Analyse. Ein Beispiel für Probleme der Merkmalsdefinition*. Aachen: Shaker.
- Meyerhoff, Jürgen, und Ulf Liebe. 2009. Status quo effect in choice experiments: Empirical evidence on attitudes and choice task complexity. *Land Economics* 85:515–528.
- Mood, Caren. 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.
- Nisic, Natascha, und Katrin Auspurg. 2009. Faktorieller Survey und klassische Bevölkerungsumfragen im Vergleich – Validität, Grenzen und Möglichkeiten beider Ansätze. In *Klein aber fein! Quantitative empirische Sozialforschung mit kleinen Fallzahlen*, Hrsg. Peter Kriwy und Christiane Groß, 211–235. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Ryan, Mandy, und Jenny Hughes. 1997. Using conjoint analysis to assess women's preferences for miscarriage management. *Health Economics* 6:261–273.
- Ryan, Mandy, Verity Watson, und Karen Gerard. 2008. Practical issues in conducting a discrete choice experiment. In *Using discrete choice experiments to value health and health care*, Hrsg. Mandy Ryan, Karen Gerard, und Mabel Amaya-Amaya, 73–88. Dordrecht: Springer.
- Steiner, Peter M., und Christiane Atzmüller. 2006. Experimentelle Vignettendesigns in Faktoriellen Surveys. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 58:117–146.
- Street, Deborah J., und Leonie Burgess. 2007. *The construction of optimal stated choice methods. Theory and methods*. Hoboken: Wiley.
- Swait, Joffre, und Wiktor Adamowicz. 2001. The influence of task complexity on consumer choice: A latent class model of decision strategy switching. *Journal of Consumer Research* 28:135–148.
- Telser, Harald. 2002. *Nutzenmessung im Gesundheitswesen. Die Methode der Discrete-Choice-Experimente*. Hamburg: Kovac.
- Telser, Harry, und Peter Zweifel. 2007. Validity of discrete-choice experiments evidence for health risk reduction. *Applied Economics* 39:69–78.
- Wallander, Lisa. 2009. 25 years of factorial surveys in sociology: A review. *Social Science Research* 38:505–520.
- Wittink, Dick R., Laksham Krishnamurthi, und Julia B. Nutter. 1982. Comparing derived importance weights across attributes. *Journal of Consumer Research* 8:471–474.

**Katrin Auspurg**, 1974, Dr., wissenschaftliche Assistentin am Fachbereich Geschichte und Soziologie der Universität Konstanz. Forschungsgebiete: Methoden der empirischen Sozialforschung, Soziale Ungleichheit, Familien- und Bildungssoziologie. Veröffentlichungen: Gruppenvergleiche bei Regressionen mit binären abhängigen Variablen – Probleme und Fehleinschätzungen am Beispiel von Bildungschancen im Kohortenvergleich, *Zeitschrift für Soziologie* 40, 2011 (mit T. Hinz); Herausbildung einer akademischen Elite? Zum Einfluss der Größe und Reputation von Universitäten auf Forschungsförderung (mit T. Hinz und J. Güdler), *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 60, 2008.

**Ulf Liebe**, 1978, Dr. phil., Juniorprofessor für Soziologie Ländlicher Räume an der Georg-August-Universität Göttingen und der Universität Kassel. Forschungsgebiete: Umweltsoziologie, Wirtschafts- und Organisationssoziologie, Soziologische Theorie. Veröffentlichungen: Status quo effect in choice experiments: Empirical evidence on attitudes and choice task complexity, *land economics* 85, 2009 (mit Jürgen Meyerhoff); Zahlungsbereitschaft für kollektive Umweltgüter. Soziologische und ökonomische Analysen, Wiesbaden 2007.