

## **Wo liegen die Grenzen quantitativer Längsschnittanalysen?**

Nina Baur

Bamberger Beiträge zur empirischen Sozialforschung

Herausgeber:

Gerhard Schulze und Nina Baur

Nr. 23, 2004

ISSN 1611-583X

## Bamberger Beiträge zur empirischen Sozialforschung

1	<b>Einführung in die Methoden der empirischen Sozialforschung</b>	Gerhard Schulze	2002
2	<b>Einführung in die Methoden der empirischen Sozialforschung. Übungsaufgaben und Lösungen</b>	Gerhard Schulze	2002
3	<b>Die biographische Methode. Ein Verfahren zur qualitativen Analyse individueller Verlaufsmuster in den Sozialwissenschaften (2., unveränderte Auflage)</b>	Nina Baur	2003
4-1	<b>Sozialwissenschaftliche Datenanalyse am PC für Fortgeschrittene. Ein Arbeitsbuch mit SPSS für Windows</b>	Nina Baur	2003
5	<b>Takeoff der Auswertung. Zur Vorbereitung statistischer Analysen</b>	Nina Baur	2003
8	<b>Univariate Statistik. Eine Einführung in den Umgang mit eindimensionalen Häufigkeitsverteilungen mit SPSS für Windows</b>	Nina Baur	2003
9	<b>Bivariate Statistik, Drittvariablenkontrolle und das Ordinalskalensproblem. Eine Einführung in die Kausalanalyse und in den Umgang mit zweidimensionalen Häufigkeitsverteilungen mit SPSS für Windows (2., korrigierte Auflage)</b>	Nina Baur	2003
10	<b>Multiple lineare Regressionsanalyse. Eine Einführung für Sozialwissenschaftler mit SPSS für Windows</b>	Sabine Fromm	2003
12	<b>Faktorenanalyse und Reliabilitätsanalyse. Eine Einführung für Sozialwissenschaftler mit SPSS für Windows</b>	Sabine Fromm	2003
13	<b>Wie kommt man von den Ergebnissen der Faktorenanalyse zu Dimensionsvariablen? Eine Einführung in die Dimensionsbildung mit SPSS für Windows (2., korrigierte Auflage)</b>	Nina Baur	2003
19-1	<b>Quantitative Analyse zeitlicher Veränderung Band 1: Überblick und theoretische Grundlage</b>	Nina Baur (Hg.)	2003
20	<b>Präsentation statistischer Daten. Eine Einführung für Sozialwissenschaftler mit SPSS für Windows</b>	Simone Zdrojewski / Jan D. Engelhardt	2003
21	<b>Datenaufbereitung. Arbeitsschritte zwischen Erhebung und Auswertung quantitativer Daten</b>	Detlev Lück	2003
22	<b>Arbeiten mit MAXqda. Kurze Einführung in die computergestützte Analyse qualitativer Daten</b>	Britta Wagner	2003
23	<b>Wo liegen die Grenzen quantitativer Längsschnittsanalysen?</b>	Nina Baur	2004

ISSN: 1611-583X

Herausgegeben von Gerhard Schulze und Nina Baur  
Professur für Methoden der empirischen Sozialforschung

Otto-Friedrich-Universität Bamberg

96045 Bamberg

<http://www.uni-bamberg.de/sowi/empirie/Download/bbzesf.html>

# Inhalt

<b>1</b>	<b>Einleitung</b> .....	<b>4</b>
<b>2</b>	<b>Selbstbegrenzung der Forschergemeinschaft</b> .....	<b>8</b>
2.1	Nicht-Anwendung von Längsschnittsanalysen .....	8
2.2	Multiplizierung typischer Probleme quantitativer Forschung .....	8
a)	Mangelnder Theoriebezug .....	8
b)	Modellierungsfehler und Überinterpretation der Aussagekraft von Items .....	8
c)	Messfehler durch schlechtes Fragebogendesign .....	10
d)	Überschätzung der Leistungsfähigkeit der Inferenzstatistik .....	10
2.3	Unreflektiertes Verhalten in einer spezifischen Forschungstradition .....	11
a)	Fixierung auf die standardisierte Befragung .....	11
b)	Fixierung auf Individualdaten .....	12
c)	Fixierung auf bestimmte Auswertungsverfahren.....	12
<b>3</b>	<b>Grenzen der bisherigen Auswertungsverfahren</b> .....	<b>13</b>
<b>4</b>	<b>Begrenzung durch die Datengrundlage</b> .....	<b>14</b>
4.1	Spezielle Stichprobenprobleme bei quantitativen Längsschnittsanalysen .....	14
a)	Vergrößerung des Konfidenzintervalls .....	14
b)	Lücken in der Protokollierung .....	14
c)	Wandel der Grundgesamtheit.....	16
4.2	Mangel an Daten.....	17
a)	Wandel der Forschungspraxis.....	17
b)	Unterschiedliche Zeitfenster .....	17
c)	Wandel des Kontextes .....	18
<b>5</b>	<b>Theoretische Grenzen</b> .....	<b>21</b>
5.1	Neue Forschungsfragen .....	21
5.2	Komplexe Modelle .....	21
5.3	Deutungs- und Interaktionsmuster .....	21
<b>6</b>	<b>Schlussfolgerungen für die Forschungspraxis</b> .....	<b>22</b>
6.1	Optimierung des Prozesses quantitativer Sozialforschung .....	22
a)	Erweiterung des gängigen Modells des Forschungsprozesses um zusätzliche Phasen .....	22
b)	Datenerhebung und -aufbereitung .....	22
c)	Datenauswertung .....	24
6.2	Validierung quantitativer durch qualitative Verfahren in bestimmten Forschungsphasen (Phasenmodell qualitativer und quantitativer Forschung) .....	24
a)	Hypothesengenerierung .....	25
b)	Wandel der Grundgesamtheit.....	25
c)	Produktionsbias und Selektionsbias.....	25
d)	Datenerhebung und –aufbereitung.....	25
6.3	Ergänzung quantitativer durch qualitative Forschungsergebnisse (Komplementaritätsmodell qualitativer und quantitativer Forschung) .....	25
<b>7</b>	<b>Literatur</b> .....	<b>26</b>

# 1 Einleitung<sup>1</sup>

Die Vorzüge quantitativer Längsschnittsanalysen sind unbestritten: Gegenüber Querschnittsanalysen sind vergleichbare Längsschnittsanalysen *immer* vorzuziehen, denn Soziologen interessieren sich für Regelmäßigkeiten im Handeln und Denken von Menschen und Gruppen von Menschen (*Schulze (2002a); Schulze (2003a); Schulze (2003b)*). Diese Regelmäßigkeiten manifestieren sich erst in der Zeit. Nur Längsschnittsverfahren ermöglichen die Analyse von Wandlungsmustern, die Analyse komplexer Kausalzusammenhänge sowie eine effizientere Drittvariablenkontrolle (*Blossfeld / Müller (1996); Blossfeld / Rohwer (1996); Blossfeld / Klijzing / Pohl / Rohwer (1996); Blossfeld / Prein (1998); Blossfeld / Steinhage (1999)*).<sup>2</sup>

Quantitative Längsschnittsanalysen sind aber nicht nur von quantitativen Querschnittsanalysen sondern auch von qualitativen Längsschnittsanalysen abzugrenzen. Im Gegensatz zu qualitativen Verfahren ermöglichen quantitative Längsschnittsverfahren zu beschreiben, wie sich Verteilungen individueller Merkmale in Kollektiven wandeln (*Schulze (2002a)*). Diese Informationen komprimieren sie auf einfache Modelle mit wenigen Parametern. Dies erleichtert die Kommunikation über diese Phänomene. Die Klarheit der Modelle erlaubt es, sie miteinander und mit anderen Phänomenen zu verknüpfen und so die Theoriebildung voranzutreiben. Die Stichproben sind im Normalfall wesentlich größer als die qualitativen Analysen. Handelt es sich um repräsentative Zufallsstichproben, ist es wesentlich wahrscheinlicher als bei qualitativen Verfahren, dass die Modelle verallgemeinert werden können. Oft erlauben die Modelle sogar sehr präzise Prognosen.

Quantitative Längsschnittsanalysen erlauben also, einige (aber nicht alle!) wesentliche sozialwissenschaftliche Erkenntnisinteressen besser zu verfolgen als andere Verfahren. Deshalb halte ich quantitative Längsschnittsanalysen in den Sozialwissenschaften nicht nur für sinnvoll sondern auch für notwendig. Trotzdem oder gerade deshalb halte ich es für ebenso notwendig, die Schwächen quantitativer Längsschnittsanalysen herauszuarbeiten, um empirische Ergebnisse nicht falsch oder überzuiinterpretieren. Diesen Grenzen möchte ich mich im folgenden widmen.

Zu diesem Zweck möchte ich zunächst den Begriff der „quantitativen Längsschnittsanalyse“ präzisieren. In der obigen Diskussion habe ich bereits implizit vorweggenommen, dass sich quantitative Längsschnittsanalysen zunächst von anderen Verfahren unterscheiden (vgl. *Grafik 1*):

- 1) *Längsschnittsanalysen vs. Querschnittsanalysen*: Quantitative Längsschnittsanalysen unterscheiden sich von quantitativen Querschnittsanalysen durch den Einbezug mehrerer Zeitpunkte in die Analyse derselben Grundgesamtheit. Wie bereits erwähnt, sind Längsschnittsanalysen Querschnittsanalysen *immer* vorzuziehen.
- 2) *Quantitative Verfahren vs. Qualitative Verfahren*: Quantitative Längsschnittsanalysen sind des weiteren von qualitativen Längsschnittsanalysen abzugrenzen.<sup>3</sup> Darüber, worin der Unterschied zwischen „qualitativen“

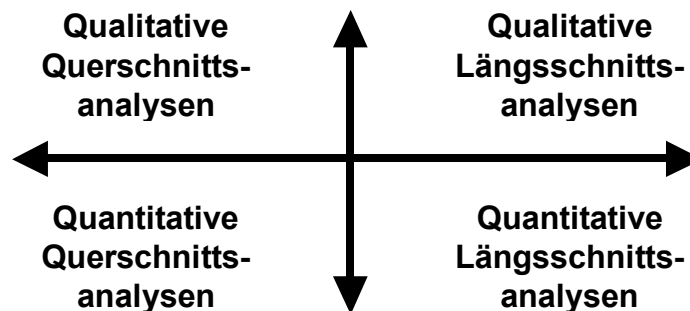
---

<sup>1</sup> Dieser Text ist eine leicht überarbeitete Fassung meines Aufnahmenvortrages bei der Sektion „Methoden der empirischen Sozialforschung“ der DGS auf der Tagung „Design und Analyse von Längsschnitten“ am 27. / 28.02.2004 in Mannheim. Ich danke meinen Bamberger Kollegen für Hinweise zur Verbesserung dieses Textes. Insbesondere danke ich Daniela Grunow, Christian Lahusen, Joachim Behnke, Nathalie Behnke, Sabine Fromm, Karin Kurz, Detlev Lück, Sandra Buchholz, Monika Taddicken, Matthias König, Christoph Wunder, Dana Hamplova, Štěpánka Pollnerová, Hanspeter Blossfeld und Gerhard Krug.

<sup>2</sup> Ein Beispiel hierfür ist der Vergleich der Regressionsanalyse mit der Zeitreihenanalyse (als Erweiterung der Regressionsanalyse um eine Zeitkomponente): Die Regressionsanalyse soll komplexe Kausalverhältnisse klären. Meist werden zu diesem Zweck Querschnittsdaten analysiert. Ein wesentlicher Aspekt der Kausalität ist aber, dass die Ursache der Wirkung vorausgegangen ist. Man benötigt also mindestens zwei Zeitpunkte, um Kausalität feststellen zu können. Geht es um Zyklen, also um sich über längere Zeit immer wiederholende Kausalitätsbeziehungen, sind sogar mehrere Messzeitpunkte nötig. Die Regressionsanalyse über Querschnittsdaten kann also gar nicht klären, was sie zu klären beansprucht – im Gegensatz zur Zeitreihenanalyse. Ist die Anwendung der Regressionsanalyse damit unsinnig? In den meisten Fällen, ja. Wie ich weiter unten noch erläutern werde, existieren aber Ausnahmen: Manche Variablen verändern sich im Zeitverlauf nur sehr langsam oder gar nicht. Beispiele sind das Geschlecht, der Bildungsgrad und fest verankerte Wertmuster. Bei anderen Variablen weiß man, wie sie sich im Zeitverlauf geändert haben. Menschen werden beispielsweise immer älter. Wenn ich weiß, wie alt Sie heute sind, weiß ich, wie alt Sie vor einem Jahr waren. Dasselbe gilt für Variablen wie Berufserfahrung, Jahre seit der Eheschließung usw. Manche Variablen beeinflussen sich fast zeitgleich gegenseitig. Den Einfluss solcher Variablen auf die abhängige Variable kann man auch bei Querschnittsdaten berechnen. Allerdings bleibt immer ein Rest an Unsicherheit, ob sie sich nicht doch ändern. Deshalb ist die Zeitreihenanalyse immer der Regressionsanalyse vorzuziehen, sofern entsprechende Daten vorliegen.

<sup>3</sup> Bisweilen wird den qualitativen und quantitativen noch eine dritte Kategorie gegenübergestellt: die historischen Verfahren. Eine genauere Analyse der Methodenliteratur und empirischen Forschung in der Geschichtswissenschaft zeigt aber, dass Historiker vor denselben Problemen der Datenanalyse stehen wie Soziologen und die selben Datenanalyseverfahren verwenden – also auch qualitative und quantitative Verfahren. Die Sozialwissenschaften können von der Geschichtswissenschaft eher im Bereich der Datenerhebung und -aufbereitung lernen. Insofern ist es unsinnig, historische von sozialwissenschaftlichen Methoden abzugrenzen. Beide haben denselben Gegenstand, wie z. B. *Marc Blochs* klas-

Grafik 1: Typen von Analyseverfahren



und „quantitativen“ Verfahren liegt, existieren verschiedene Auffassungen (vgl. hierzu *Behnke / Behnke / Baur* (2004)). Um Missverständnissen vorzubeugen, definiere ich deshalb „qualitative Sozialforschung“ als eine typische Kombination aus einem spiralförmigen Forschungsprozess sowie kleinen, bewusst ausgewählten Stichproben; einer großen Zahl von Variablen, die offen erhoben, transkribiert, – bei Bedarf – in CAQDAS eingelesen und mit Hilfe von regelgeleiteten Verfahren ausgewertet werden, die der Fragestellung und der offenen Struktur der Daten angemessen sind. Mit „quantitativer Sozialforschung“ bezeichne ich eine typische Kombination aus linearem Forschungsprozess sowie großen, möglichst nach dem Zufallsprinzip ausgewählten, zumindest aber repräsentativen Stichproben einer vergleichsweise geringen Zahl von Variablen, die geschlossen erhoben, in einen Datensatz eingelesen und mit Hilfe statistischer Verfahren ausgewertet werden. Es handelt sich hierbei nur um idealtypische Kombinationen, und die Übergänge zwischen den Verfahren sind fließend. In einzelnen Punkten kann ein konkreter Forschungsprozess von den Idealtypen abweichen. Beide Forschungsstrategien sind gleichermaßen strukturiert und planmäßig. Um bereits auf eines meiner Ergebnisse vorweg zu greifen: Ich sehe die Schwächen quantitativer Längsschnittsanalysen vor allem in diesem Bereich.

Ich grenze quantitative Längsschnittsanalysen nicht nur gegenüber anderen Verfahren ab sondern unterteile diese selbst nochmals in Unterkategorien. Dabei unterscheide ich zwischen Erhebungsdesigns und Auswertungsverfahren (vgl. *Grafik 2*; für eine detaillierte Beschreibung siehe auch *Baur* (2003a)).

Um Zeitvergleiche vornehmen zu können, sind bestimmte Erhebungsdesigns erforderlich. Geeignet sind insbesondere die folgenden sechs Designtypen, die in den Sozialwissenschaften allerdings unterschiedlich häufig angewandt werden.<sup>4</sup> Zunächst zu nennen ist das Trenddesign. Beispiele sind ALLBUS und ISSP. Ein Beispiel für zu wissenschaftlichen Zwecken im Paneldesign erhobene Daten ist das Sozio-ökonomische Panel (SOEP), ein Beispiel aus der privaten Forschung sind Panels aus der Marktforschung. Eine Abart des Panels sind prozessgenerierte quantitative Daten. Diese entstehen durch laufende Protokollierung. Beispiele sind beim Besuch bestimmter Internetseiten mitprotokollierte Benutzerinformationen, Kundendatenbanken von Unternehmen, Aktenvermerke und regelmäßig ausgefüllte Formulare in Bürokratien, aber auch von Forschern bewusst angelegte Protokolle.

So bizarr es zunächst klingen mag – auch Querschnittsdesigns lassen sich für Längsschnittsanalysen nutzbar machen, und zwar auf zweierlei Weise: Eine Möglichkeit ist die Retrospektivbefragung, mit deren Hilfe Befragte nachträglich Ereignisse in der Vergangenheit aus ihrer Erinnerung rekonstruieren. Beispielsweise erzählen Befragte bei der Erstbefragung der German Life History Study (GLHS) ihre bisherige Biographie bis zum Erhebungszeitpunkt. Es handelt sich also um ein Querschnittsdesign, mit dessen Hilfe Längsschnittsdaten erhoben werden. Aber auch wenn in Querschnittserhebungen keinerlei Items vergangene Ereignisse messen, lassen sich diese Erhebungen später mittels Replikationsstudien für Zeitvergleiche nutzbar machen. Die erste Einmalbefragung stellt zwar nur eine Momentaufnahme ohne zeitlichen Bezug dar. Übernimmt ein Forscher jedoch später forschungsrelevante Fragen in ihrer *exakten* Formulierung in einen eigenen Fragebogen, kann er Veränderungen zwischen den Befragungszeitpunkten feststellen.<sup>5</sup>

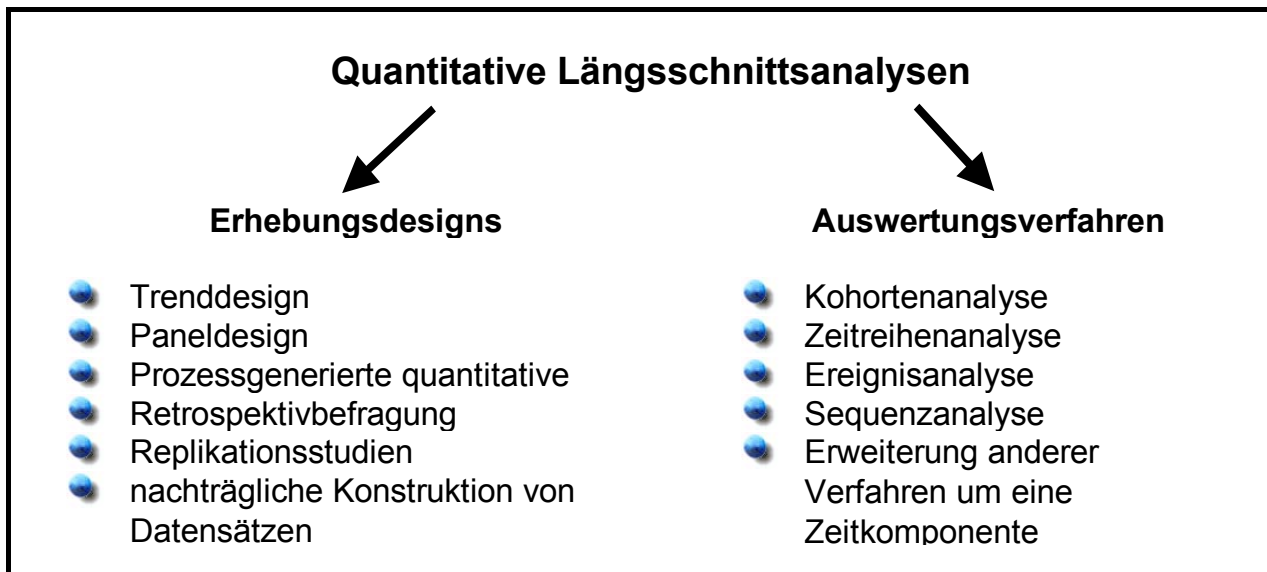
---

sische Definition der Geschichtswissenschaft zeigt: Geschichtswissenschaft ist die „Wissenschaft von den Menschen ... in der Zeit“ (*Bloch* (2002): 32). Diese Definition könnte auch von einem Soziologen stammen.

<sup>4</sup> Jeder dieser Designtypen hat eigene, designtypische Probleme. Diese Erhebungsdesigns erläutere ich ausführlich in *Baur* (2003a). Zusätzliche Erläuterungen finden Sie auch in *Schulze* (2003b) und *Blossfeld, Hamerle und Mayer* (1986).

<sup>5</sup> Hierzu ein Beispiel: *Thomas Müller-Schneider* (1994, 1996, 2000) zeigt, dass für solche Zeitvergleiche nicht einmal Replikationsstudien notwendig sind. Sein theoretischer Ausgangspunkt waren zwei konkurrierende Modelle sozialer Ungleichheit: das Schichtmodell (*Bolte* (1961); *Bolte / Hradil* (1988); *Daheim* (1961); *Daheim / Scheuch* (1961); *Hra-*

Grafik 2: Typen quantitativer Längsschnittsanalysen



Eine letzte Möglichkeit, Daten für Querschnittsanalysen zu erheben, ist, die nachträgliche Konstruktion von Datensätzen mit Hilfe qualitativer prozessgenerierter Daten. Dies ist die gängige Methode der quantitativen Geschichtswissenschaft. Ein frühes Beispiel der „histoire serielle“ (*Braudel* (1950, 1963)) ist *Fernand Braudels* Geschichte der Preise (*Braudel* (1967)). Doch auch in der Soziologie werden vereinzelt Datensätze nachträglich rekonstruiert. Ein Beispiel ist die *Paul Windolf* und Kollegen aufgebaute Unternehmensdatenbank: Mit Hilfe verschiedener Datenquellen, darunter Geschäftsberichte und Finanzdaten, bauten sie einen Datensatz auf, mit dessen Hilfe sie den Wandel der Unternehmenskontrolle in Deutschland von 1992 bis 2000 sowie in Frankreich von 1996 bis 2000 nachzeichnen konnten (*Beyer und Windolf* (1995); *Windolf* (2002)). *Windolf* (2002) zeigt mit Hilfe dieser Daten, dass der Rheinische Kapitalismus derzeit im Umbruch begriffen ist und sich an das angelsächsische Unternehmenskontrollsystem annähert. In seinem aktuellen Forschungsprojekt baut *Windolf* (o. J.) einen vergleichbaren Datensatz für die Jahre 1896, 1914, 1924, 1932 und 1938 auf (und erweitert die Datenbasis um die USA).

Ein zweites Beispiel ist eines meiner eigenen Forschungsprojekte, das ich in enger Kooperation mit den Mitarbeitern des Projekts „*unempol*“ (*Giugni / Statham* (2002)), insbesondere *Christian Lahusen* durchführe: Unser Forschungsinteresse ist der Wandel des Diskurses über die Arbeitslosigkeit. Hierzu kodierten wir mit Hilfe eines fast 70seitigen Codebuchs Claims über Arbeitslosigkeit, die in Zeitungsartikeln gemacht wurden. Claims sind verbale und nonverbale strategische Interventionen, die ein politischer Akteur zu einem Thema in der Öffentlichkeit macht (*Giugni / Statham* (2002)). Wir arbeiteten dabei mit zwei verschiedenen Zeitrahmen: In meinem Projekt wurden von 1964 bis 2000 alle vier Jahre Artikel aus möglichst unterschiedlichen Zeitungen kodiert, die Projektmitarbeiter von „*unempol*“ erfassten von 1995 bis 2002 für jeden dritten Tag Zeitungsartikel aus der Süddeutschen.

*dil* (1999); *Rosenmayr* (1961)) und das Erlebnismodell (*Schulze* (1996)). Das eine geht von hierarchischen Strukturen aus, das andere von subjektiver Selbstzuschreibung. Beide Modelle sind empirisch fundiert, allerdings stützen sie sich auf unterschiedliche Daten, die zudem noch im Abstand von dreißig Jahren erhoben wurden. *Müller-Schneider* wollte überprüfen, ob das Schichtmodell oder *Schulzes* (1996) Milieumodell besser geeignet ist, um soziale Ungleichheit in Deutschland zu modellieren. Hierzu grub er sieben zeitlich gestaffelte Standardrepräsentativumfragen aus, die zumindest zu einem Teil ähnliche, forschungsrelevante Fragen enthalten. Zu diesen Daten gehören die Leseranalysen 1953/54, 1963, 1970 und 1976/77. Seit 1986 enthalten ALLBUS und das Wohlfahrtssurvey Fragen, die nach einigen Modifikationen erlauben, Alltagsorientierungen einer Person zu messen, so dass *Müller-Schneider* drei weitere folgende Messzeitpunkte hinzuziehen kann: 1985/86, 1993 und 1998. Mit Hilfe dieser Daten konnte *Müller-Schneider* zeigen, dass das Schichtmodell langsam in das Milieumodell übergeht.

Diese Daten weisen einige Probleme auf: Die Grundgesamtheiten sind nicht exakt gleich. Einige der erhobenen Merkmale waren zwar unmittelbar vergleichbar. Andere dagegen musste *Müller-Schneider* modifizieren, um die Vergleichbarkeit sicherzustellen. Die beiden Grundprobleme – systematische Fehler bei der Stichprobenziehung und geringe Anzahl kommensurabler Items – bleiben. Sie sind bei der Interpretation der Ergebnisse zu beachten. *Müller-Schneider* geht dieses Fehlerrisiko ein, weil keine anderen Daten verfügbar sind und das Fehlerrisiko relativ gering scheint (*Müller-Schneider* (1996): 192-195; *Müller-Schneider* (2000): 362-364).

Vom Erhebungsdesign zu unterscheiden sind Verfahren der Datenanalyse.<sup>6</sup> Verfahren, die spezifisch zur Analyse von Längsschnittdaten geeignet sind, lassen sich grob in fünf Gruppen aufteilen:

Das wohl bekannteste Beispiel für die Kohortenanalyse (Glenn (1977); Huinink (1988); Huinink / Mayer (1990) sind Ingleharts Wertewandelstudien (Inglehart (1989, 1997); Abramson / Inglehart (1995)). Ein Beispiel für die Zeitreihenanalyse (Hartung / Elpelt / Klösener (2002); Schlittgen / Streitberg (2001)) ist die oben erwähnte Studie Braudels zur Geschichte der Preise. Ein anderes Beispiel ist die Forschung über Wirtschaftszyklen bzw. Wellen im Wirtschaftsprozess, wie sie von Wirtschaftswissenschaftlern und Postfordismus-Theoretikern diskutiert werden. Beispiele für die Ereignisanalyse (auch „Sterbetafelanalyse“, „Überlebensanalyse“, „event history analysis“, „survival analysis“, Andreß (1985); Diekmann (1988); Blossfeld / Rohwer (2002)) sind die Arbeiten in den Projekten von Hans-Peter Blossfeld, z. B. derzeit im Globalife-Projekt in Bamberg. Abbott und Tsay (2000) nennen einige Beispiele für die Sequenzanalyse und Optimal Matching-Analyse (Abbott (1995), Fromm (2003), Erzberger (2001), Erzberger / Prein (1997)). Eine letzte Möglichkeit ist die Erweiterung anderer quantitativer Verfahren um eine zeitliche Komponente. Beispielsweise erweitert Thomas Müller-Schneider (1994, 1996, 2000) die Korrespondenzanalyse um eine zeitliche Komponente. Ich will ein weiteres Beispiel aus dem Globalife-Projekt nennen: Daniela Grunow (2004) untersuchte den Wandel der Erwerbskarrieren von dänischen Männern in den mittleren Lebensjahren. Die Datenbasis waren dänische Registerdaten. Diese Daten geben nur den jährlichen Stand der Erwerbskarriere wieder – eine Ereignisanalyse schied also aus. Also wählte sie eine um eine Zeitkomponente erweiterte Version der logistischen Regressionsanalyse.

Spezifische Erhebungsdesigns implizieren *nicht* bestimmte Auswertungsverfahren, und bestimmte Auswertungsverfahren bedingen nicht spezifische Erhebungsdesigns. Dennoch gehen bestimmte Auswertungsverfahren und Erhebungsdesigns zum Teil Hand in Hand. Der Grund sind die Voraussetzungen für die Anwendung eines bestimmten Verfahrens. Zeitreihenanalysen setzen metrische Daten voraus, die Ereignisanalyse Ereignisdaten. Letztere können nur über Retrospektivbefragungen oder laufende Protokollierung gewonnen werden.

Zusammenfassend sind also quantitative Längsschnittsverfahren einerseits von qualitativen und Querschnittsanalysen zu unterscheiden, zum anderen muss der Analyseprozess differenziert werden in zeitbezogene quantitative Erhebungsdesigns und Auswertungsverfahren. Dieses Raster verwende ich als Grundlage für die folgende Diskussion der Grenzen quantitativer Längsschnittsanalysen. Ich habe die Grenzen quantitativer Längsschnittsanalysen in vier grobe Bereiche unterteilt, die allerdings nicht trennscharf sind: Der erste Bereich ist die Selbstbegrenzung der Forschergemeinschaft. Dies ist keine grundsätzliche Grenze quantitativer Längsschnittsanalysen sondern auf die bisherige Forschungspraxis zurückzuführen. Sie lässt sich also relativ leicht beheben. Dasselbe gilt für den zweiten Bereich – die Grenzen der bisherigen Auswertungsverfahren. Durch Weiterentwicklung der bisherigen Verfahren ließe sich dieses Problem beseitigen. Schwieriger ist dies beim dritten Bereich, der Begrenzung durch die Datengrundlage. Teilweise könnte dieses Problem überwunden werden, indem auf andere Erhebungsdesigns zurückgegriffen wird. Bestimmte Informationen sind allerdings nicht rekonstruierbar, also für immer verloren. Der letzte Bereich, den ich anspreche, sind theoretische Grenzen quantitativer Verfahren. Diese Grenzen sind innerhalb der quantitativen Sozialforschung nicht zu lösen. In diesem Bereich ist also eine Kombination qualitativer und quantitativer Verfahren erforderlich.

---

<sup>6</sup> Jedes dieser Verfahren hat ein spezifisches Erkenntnisinteresse, eigene Voraussetzungen an die Daten, die es verarbeitet, sowie ein besonderes theoretisches Potential. Ich beschreibe sie kurz in Baur (2003a). Trotz dieser Unterschiede ist der obige Kategorisierungsversuch statistischer Verfahren zur Längsschnittsanalyse nicht trennscharf. Er richtet sich vielmehr nach häufig angewandten Auswertungsverfahren. Statistisch gesehen, sind die meisten Auswertungsverfahren miteinander verwandt: Beispielsweise ist die Ereignisanalyse ein Spezialfall der Zeitreihenanalyse, die wiederum ein Spezialfall der Regressionsanalyse ist. Diese wiederum ist ein Spezialfall Allgemeiner Linearer Modelle. Die Kohortenanalyse ist ein Spezialfall der Kontingenzanalyse. Werden multivariate Querschnittsverfahren um eine Zeitkomponente erweitert, bedeutet dies in der Regel, dass sie mit der Zeitreihenanalyse verknüpft werden.

## 2 Selbstbegrenzung der Forschergemeinschaft

Die erste Grenze quantitativer Längsschnittsanalysen könnte man als „Selbstbegrenzung der Forschergemeinschaft“ bezeichnen. Gemeint ist damit, dass es sich um kein grundsätzliches Problem sondern ein Problem der Forschungspraxis handelt. Damit lässt es sich prinzipiell leicht beheben. Die Probleme der Forschungspraxis habe ich in drei Problembündel zusammengefasst: die Nicht-Anwendung von Längsschnittsanalysen, die Multiplizierung typischer Probleme quantitativer Forschung sowie unreflektiertes Verhalten in einer spezifischen Forschungstradition.

### 2.1 Nicht-Anwendung von Längsschnittsanalysen

Forscher beschränken die Leistungsfähigkeit ihrer Analysen in der Praxis häufig dadurch, dass sie Auswertungsverfahren zur Längsschnittsanalyse schlicht nicht anwenden. Nach wie vor beschränken sich die meisten empirischen Studien auf Querschnittsanalysen (*Blossfeld / Steinhage* (1999); *Engel* (2002)). Wie problematisch dies ist, realisieren Forscher häufig nicht aufgrund eines typischen Denkfehlers: Sie verwechseln statistische Zusammenhänge mit Kausalzusammenhängen. Statistische Zusammenhänge können aber auch zustande kommen, wenn es sich um einen dimensional Zusammenhang handelt. Ebenso können Drittvariablen zu einem statistischen Zusammenhang führen (*Mayer* (1998); *Schulze* (2002)). Doch selbst diejenigen Forscher, die Längsschnittsanalysen betreiben, gehen häufig nicht weit genug in die Zeit zurück, um ihre Forschungsfrage zu beantworten – die meisten Längsschnittsanalysen behandeln Zeiträume ab den 1970ern oder 1980ern. Dies ist, was *Norbert Elias* als „Rückzug der Soziologie in die Gegenwart“ bezeichnete.

### 2.2 Multiplizierung typischer Probleme quantitativer Forschung

Da es sich bei den oben genannten Erhebungsdesigns um Forschungsdesigns für standardisierte Erhebungen handelt, gelten banalerweise auch die Einschränkungen, die für quantitative Verfahren im allgemeinen gelten. Dies ist also kein spezifisches Problem von quantitativen Längsschnittanalysen sondern von quantitativen Analysen allgemein. Da mehrere Zeitpunkte betrachtet werden, multiplizieren sich Schwächen im Forschungsdesign leicht. Diese Mängel hier anzuführen, erscheint insofern banal, da sie bereits oft diskutiert wurden. Allerdings werden viele Probleme nach wie vor nicht einmal in Lehrbüchern zur empirischen Sozialforschung thematisiert (vgl. hierzu *Ludwig Mayerhofer* (2003)). Dass diese Probleme in der Forschungspraxis häufig ignoriert werden, zeigt des weiteren, dass diese Mängel nie oft genug angeführt werden können. In der derzeitigen Forschungspraxis sehe ich vor allem folgende Probleme:

#### a) Mangelnder Theoriebezug

Selbstverständlich treiben quantitative Sozialforscher in vielen Fällen mit Hilfe ihrer empirischen Analysen die Theoriebildung voran. Einige Beispiele habe ich bereits genannt. Ebenso wie manche soziologischen Theoretiker die Realität völlig ausblenden, betreiben aber auch manche quantitative Sozialforscher theorieleose Forschung. Zahlreiche Analysen erwecken den Eindruck, dass Daten und Methode nicht aufgrund bestimmter theoretischer Erwägungen ausgewählt wurden, sondern dass die Daten analysiert wurden, weil der Forscher zufällig Zugang zu ihnen hatte. Die Forschungsfrage wird an die Daten angepasst (*Mayer* (1998)).<sup>7</sup>

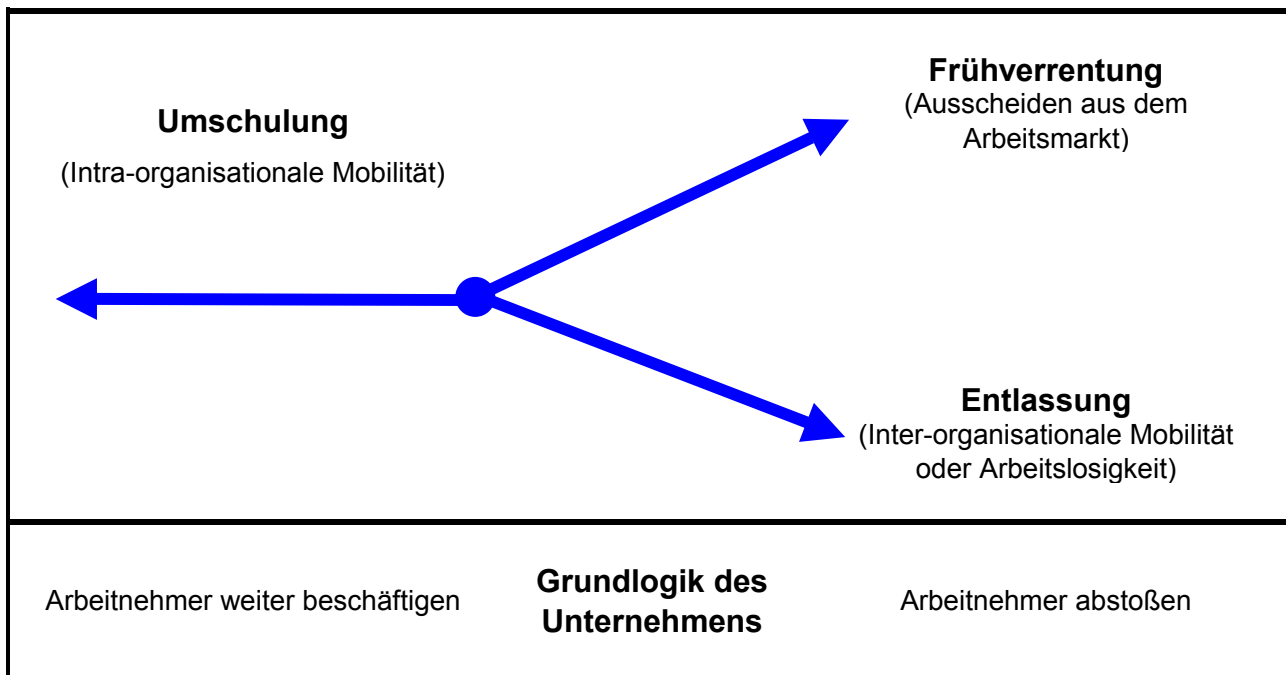
#### b) Modellierungsfehler und Überinterpretation der Aussagekraft von Items

Ein verwandtes Problem ist, dass bei der Modellierung Fehler begangen werden und die Aussagekraft von Items häufig überinterpretiert wird. Bisweilen zielen die analysierten Daten einfach am Erkenntnisinteresse vorbei. Ein Beispiel *Gerhards* (2002a, 2002b) Versuch, die Reputation der deutschen Soziologen zu messen.<sup>8</sup>

<sup>7</sup> Um sie zu umgehen, müssen die Forscher deutlich mehr theoretische Vorarbeit leisten, bessere Theorien ausarbeiten, ihre Modelle genauer spezifizieren, Längsschnittdaten verwenden, mit Hilfe interkultureller Vergleiche nach Konstanten suchen, die Grenzen ihrer Daten und Verfahren (aner)kennen und Unschärfe akzeptieren (*Mayer* (1998)).

<sup>8</sup> Zur Kritik an *Gerhards* siehe *Allmendinger* (2002); *Rammert* (2002); *Endruweit* (2002); *Schulz-Schaeffer* (2002); *Endruweit* (2003) und *Klüver* (2003)). Diese bezog sich aber bis auf einige Ausnahmen auf die Auswahl der Zeitschriften, die er heranzog, um die Reputation einzelner Fachvertreter zu messen, nicht auf die Auswahl der Kriterien zur Messung von Reputation oder gar die Frage, ob man Reputation überhaupt wissenschaftlicher Leistung und diese wiederum Menge des Outputs entspricht.



**Grafik 3: Bezugsrahmen für die Analyse des Übergangs in den Ruhestand**

(Quelle: Buchholz und Hofäcker (2003), eigene Übersetzung)

Ebenso ignorieren Forscher, dass Fragen nur aus ihrem Kontext verstanden werden können und dass statistische Modelle je nach Kontext unterschiedlich sein können und müssen. Bei Querschnittsanalysen der eigenen Gesellschaft fällt dieses Problem selten auf, da wir den Kontext kennen. Um diese Problematik zu verdeutlichen, führe ich hier ein gelungenes Beispiel an, wie man dieses Problem umgehen kann: Buchholz und Hofäcker (2003) Überlegungen zu den Erwerbskarrieren älterer Männer:

Ein heute lebender Deutscher würde intuitiv annehmen: „Naja, die gehen halt irgendwann Mitte 50 und Mitte 60 in den Ruhestand.“ Buchholz und Hofäcker (2003) zeigen, dass dies nicht so einfach ist (vgl. Grafik 3): Aus Unternehmenssicht sind ältere Arbeitnehmer häufiger weniger belastbar und verfügen eher über veraltetes Wissen. Das Unternehmen kann sich nun entscheiden, ob es diese Arbeitnehmer weiterbeschäftigt oder freisetzt. Will es sie weiterbeschäftigen, müssen die Arbeitnehmer umgeschult und möglicherweise innerhalb des Unternehmens anders eingesetzt werden. Entschließt sich das Unternehmen, die Arbeitnehmer freizusetzen, kann es sie entweder frühverrenten oder entlassen. In ersterem Fall scheiden die Arbeitnehmer vom Arbeitsmarkt aus, im zweiten werden sie arbeitslos oder suchen sich einen anderen Arbeitgeber.

Diese drei Möglichkeiten sind aber nicht alle gleich wahrscheinlich: Der institutionelle Rahmen erleichtert, erschwert oder verbietet Unternehmen sogar spezifische Lösungen. Beispielsweise können in Deutschland ältere Arbeitnehmer mit langjähriger Berufserfahrung nicht so leicht entlassen werden, weshalb Unternehmen meist entweder die Frühverrentung wählen oder den betroffenen Arbeitnehmern ein Optionspaket anbieten, mit dessen Hilfe sie eine kurze Arbeitslosenphase überbrücken und danach regulär in Rente gehen. Diese Arbeitnehmer scheiden also aus dem Arbeitsmarkt aus (Arnds / Bonin (2002)). In den USA dagegen treten sehr unterschiedliche Muster auf. Unter anderem sind dort langjährige Abstiegskarrieren zu beobachten: Arbeitnehmer wechseln bis ins hohe Alter immer wieder den Job. Jeder neue Arbeitsplatz ist dabei schlechter als der vorhergehende (Buchholz / Hofäcker (2003)).

Der Kontext ist nicht nur interkulturell unterschiedlich, sondern wandelt sich auch über die Zeit. Abstiegskarrieren wie in den USA waren im letzten Jahrhundert auch für Deutschland typisch – Altersarmut zu vermeiden, war einer der Hauptgründe für die Einführung der Bismarck'schen Sozialversicherungssysteme. Um der Variation des Kontextes gerecht zu werden, müssen nicht nur für unterschiedliche Länder sondern auch für unterschiedliche Zeiten völlig unterschiedliche statistische Modelle gerechnet werden. Während Buchholz und Hofäcker (2003) diesen Kontext explizit vor der Modellbildung analysieren, fällt manchen anderen Forschern gar nicht auf, dass diese Kontextinformationen notwendig sind, weil sie nur ihre eigene Gesellschaft analysieren. Hier schöpfen sie aus ihrem Alltagswissen. Bei interkulturellen oder Zeitvergleichen greift das Alltagswissen aber nicht mehr – das Kontextwissen muss aus qualitativ-historischen Analysen stammen.

### c) Messfehler durch schlechtes Fragebogendesign

In jedem Methodenbuch ist nachzulesen, dass schlechtes Fragebogendesign Messfehler verursacht. Dennoch ist gutes Fragebogendesign alles andere als eine Selbstverständlichkeit. Selbst wenn der Fragebogen als Ganzes sorgfältig konstruiert sind, können sich Fehler in einzelnen Fragen einschleichen. Ob und wie gut einzelne Items das Erkenntnisinteresse greifen, muss mit Hilfe qualitativer Vorstudien (insbesondere semantischer Analysen) bestimmt werden.

### d) Überschätzung der Leistungsfähigkeit der Inferenzstatistik

Ein weiteres großes Problem der Forschungspraxis ist, dass Forscher regelmäßig die Leistungsfähigkeit der Inferenzstatistik überschätzen, und zwar in dreierlei Hinsicht: Erstens sind die meisten quantitativen Erhebungsdesigns von vorneherein gar nicht auf Zufallsstichproben angelegt.<sup>9</sup> Zu der Selektivität des Auswahlverfahrens kommen Ausfälle und fehlende Werte. Entsprechend ist praktisch jede quantitative Erhebung systematisch verzerrt. Damit ist im besten Fall zweifelhaft, ob der Einsatz der schließenden Statistik Sinn macht (*Gigerenzer* (1999); *Mayer* (1998)). Die „harte Linie“ wäre, die Inferenzstatistik gar nicht anzuwenden. Die „weiche Linie“ wäre anzunehmen, dass sich Zufallsfehler und systematischer Fehler überlagern. In beiden Fällen ist allerdings *auf jeden Fall* eine Abschätzung des Bias notwendig. Dies muss mit Hilfe qualitativer Methoden erfolgen, denn Gewichtungungsverfahren nutzen nicht viel: Informationen, die man nicht hat, kann man auch nicht in den Datensatz einrechnen (*Baur / Behnke / Behnke* (2004)). Nehmen wir das Beispiel der Online-Erhebungen: Selbst wenn tatsächlich einige achtzigjährigen Frauen geantwortet haben – ist es legitim anzunehmen, dass diese Achtzigjährigen repräsentativ sind für alle Achtzigjährigen, die keinen Internetanschluss haben?

Mindestens ebenso gravierend ist, dass Testverfahren häufig falsch angewandt und interpretiert werden. Hier summieren sich gleich mehrere Fehler: Die meisten Forscher wenden mechanisch „etablierte Testverfahren“ an, ohne zu fragen, ob dieses Testverfahren überhaupt sinnvoll sind (*Gigerenzer* (1999)).<sup>10</sup> Fast nie wird das Gesamtmodell getestet, sondern immer nur einzelne Variablen. *Beck-Bornholdt* und *Dubben* (2003a) zeigen, dass bei diesem Vorgehen wahrscheinlichkeitstheoretisch praktisch immer einzelne signifikante Ergebnisse herauskommen. Würde man komplexe Gesamtmodelle testen, wären in der Regel wesentlich höhere Fallzahlen als die üblichen  $n = 1000$  vonnöten, damit die Testergebnisse aussagekräftig wären. In der Regel werden Nullhypothesen mechanisch mit einem Signifikanzniveau von  $\alpha = 0,05$  getestet – unabhängig davon, ob

<sup>9</sup> En vogue sind zur Zeit vor allem Telefonumfragen, Online-Umfragen und Umfragen mit freiwilligem postalischem Rücklauf. Die Umfrageergebnisse werden im allgemeinen als repräsentativ für die Bevölkerung der BRD gewertet – das sind sie aber nicht. Eine Telefonumfrage kann allenfalls die Personen erfassen, die eine Telefon und eine öffentlich bekannte Telefonnummer besitzen. Obdachlose und auf ihre Privatsphäre bedachte Personen werden beispielsweise nicht erfasst. Im Internet sind nach wie vor junge Männer mit hoher Bildung aus technischen Berufen überrepräsentiert. Das Kauf- und Surfverhalten sind unterschiedlich. Die einzigen Verfahren, die am ehesten Zufallsstichproben erzeugen, sind die Meldekarteien der Einwohnermeldeämter und das Random-Route-Verfahren. Beide Verfahren werden wegen ihrer hohen Kosten selten angewandt. Bei allem Stichprobenverfahren besteht außerdem die Gefahr, dass die Interviewer die Fragebögen fälschen. Schließlich sinkt die Bereitschaft der Deutschen, an Befragungen teilzunehmen. Mitteilsame, aktive Personen sind in Befragungen überrepräsentiert. Oft versuchen Marktforschungsinstitute, Anreizsysteme zu schaffen. Dies wiederum führt dazu, Personen, die an diesen Anreizen interessiert sind, überrepräsentiert sind.

<sup>10</sup> Der Siegeszug quantitativer Sozialforschung seit den 1950ern wäre nicht möglich gewesen, hätte es nicht auch fachliche Argumente für die Überlegenheit der quantitativen gegenüber qualitativen Verfahren gegeben. Dieser Überlegenheitsanspruch begründet sich üblicherweise nicht oder nur teilweise mit der Möglichkeit der Mustererkennung und Komplexitätsreduktion sondern vor allem darauf, dass Daten durch die Kombination aus Zufallsstichprobe und Inferenzstatistik besser verallgemeinerbar sind als die qualitativer und historischer Forschung. Vorangetrieben wurde der Siegeszug quantitativer Sozialforschung durch eine kollektive Fehldeutung: In den 1940ern entwickelten Fisher einerseits und Neyman und Pearson andererseits zwei sich widersprechende Theorien des statistischen Testens. Bei Fisher-Tests stellt der Forscher eine Hypothese auf und überprüft, wie hoch die Wahrscheinlichkeit ist, dass die gegebenen Daten auftreten, wenn die Hypothese stimmt. Wird der Test verworfen, ist man sich relativ sicher, dass die Hypothese nicht stimmt – mehr nicht. Bei den üblichen Testverfahren ist die Nullhypothese, dass Zusammenhänge Null sind. Verwirft man diese Hypothese, weiß man zwar, dass der Zusammenhang wahrscheinlich nicht Null ist, aber noch lange nicht, welcher Art die Zusammenhänge sind oder wie stark sie sind (was eigentlich interessiert). Bei Neyman-Pearson-Tests werden mindestens zwei Hypothesen formuliert – die Null- und die Alternativhypothese. Basierend auf den theoretischen Verteilungen beider Hypothesen berechnet der Forscher die Power der Prüfverteilung ( $1-\beta$ ). Bei Zusammenhangsmaßen bezeichnet  $\beta$  die Wahrscheinlichkeit, dass man real existierende Zusammenhänge übersieht. Nur wenn die Power groß ist, sollte man deshalb das Signifikanzniveau  $\alpha$  berechnen.  $\alpha$  entspricht also *nicht* der Irrtumswahrscheinlichkeit. Die Testtheorien von Fisher und von Neyman und Pearson widersprechen sich. Lehrbuchautoren vermengten sie aber seit den 1950ern. Durch diese Vermengung scheinen statistische Tests wesentlich leistungsfähiger zu sein, als sie tatsächlich sind. Nach und nach prüften immer mehr Sozialforscher Hypothesen bei einem Signifikanzniveau von  $\alpha = 0,05$  durch – egal, ob dies dem Forschungsproblem angemessen war oder nicht. Indem sie die Inferenzstatistik fehldeute(n), gaben und geben sich Sozialwissenschaftler der kollektiven Illusion hin, dass die schließende Statistik sichere Verallgemeinerungen erlaube (*Gigerenzer* (1999); *Beatty / Daston / Gigerenzer / Krüger / Porter / Swijtink* (1999)).

dies sinnvoll ist. Hinzu kommt, dass meist *Fisher-Tests* als *Neyman-Pearson-Tests* fehlinterpretiert werden (*Beck-Bornholdt / Dubben (2003b), Behnke / Behnke / Baur (2004)*).<sup>11</sup>

Drittens verwechseln Forscher häufig statistische Signifikanz mit empirischer und theoretischer Relevanz (*Schulze (2002a), Baur (2003b)*).<sup>12</sup>

Um nicht missverstanden zu werden: Ich meine nicht, dass die schließende Statistik generell nutzlos ist sondern nur, dass sie wesentlich häufiger angewandt wird, als sinnvoll wäre. Eine völlige Fehlentwicklung ist, dass Zeitschriftenherausgeber häufig nur Publikationen annehmen, die quantitative Ergebnisse auf hohem Signifikanzniveau präsentieren. Berücksichtigt man aber die Grenzen der schließenden Statistik, schrumpft der Überlegenheitsanspruch der quantitativen gegenüber den qualitativen Sozialforschern deutlich.

Die hier genannten Probleme sind keinesfalls Nebenschauplätze. Es ist unumgänglich, zu reflektieren, inwiefern sie auftreten, und sie nach Möglichkeit zu vermeiden. Will man typische aber vermeidbare Mängel quantitativer Forschung beheben, muss man in der Regel auf qualitative Verfahren zurückgreifen. Beispielsweise zeigt *Hillmert (2002)*, wie standardisiert erhobene Ereignisdaten nachbearbeitet und so bereinigt werden können. Es ist also unbedingt erforderlich zu reflektieren, ob und wie die Daten und Auswertungsverfahren die Aussagekraft der Forschungsergebnisse beeinträchtigen. Die Ergebnisse dieser Analysen sind zudem in Publikationen explizit zu machen!

Häufig lassen sich diese Mängel trotz aller Bemühungen nicht vermeiden. In diesem Fall ist es keine Lösung, einfach so zu tun, als sei das nicht weiter problematisch. Im günstigsten Fall gehen die Vorteile quantitativer gegenüber qualitativen Längsschnittsverfahren verloren. Insbesondere sind dann die quantitativen Ergebnisse auch nicht besser verallgemeinerbar als qualitative – eher im Gegenteil. Das ist aber noch nicht das Worst Case-Szenario. Im schlimmsten Fall liegt nämlich ein teuer erhobener Datensatz vor, der aufwändig ausgewertet wurde. Der Datensatz ist aber schlichtweg nutzlos, die Ergebnisse falsch.

## 2.3 Unreflektiertes Verhaften in einer spezifischen Forschungstradition

Forscher beschränkten sich des weiteren selbst, indem sie einer spezifischen Forschungstradition verhaften, ungeachtet Ihres Erkenntnisinteresses. Der Grund hierfür liegt in der disziplinären Arbeitsteilung: Die Menschenwissenschaften unterteilen sich unter anderem in Soziologie, Politikwissenschaft, Wirtschaftswissenschaft, Psychologie und Geschichtswissenschaft (*Albert (1999)*). Innerhalb der Soziologie gibt es Spezialisten für soziologische Theorie, Methoden oder eine der zahlreichen speziellen Soziologien. Im Bereich der Methoden gibt es nicht nur Spezialisten für qualitative und quantitative Verfahren. Innerhalb des Bereichs quantitativer Verfahren kennen sich Spezialisten für ein spezifisches Verfahren nicht immer sehr gut mit anderen Verfahren aus.

Diese Spezialisierung erhöht zwar einerseits die Effizienz bei der Weiterentwicklung spezifischer Methoden, führt andererseits aber zu kollektiver Blindheit gegenüber anderen Ideen: Forschergemeinschaften verfolgen teilweise typische Enzwicklungspfade empirischer Sozialforschung. Sie haben sich auf bestimmte Verfahren eingeschossen, ungeachtet dessen, ob diese überhaupt dem Erkenntnisinteresse (noch) angemessen sind. Bestimmte Aspekte der sozialen Wirklichkeit blenden sie systematisch aus (*Albert (2001)*). Zu nennen sind insbesondere folgende Probleme:

### a) *Fixierung auf die standardisierte Befragung*

Obwohl die Logik quantitativer Verfahren dies nicht erfordert, ist die standardisierte Befragung die Haupterhebungsform der Sozialwissenschaften. Sozialwissenschaftler analysieren dagegen nach wie vor relativ selten prozessgenerierte Daten. In den seltenen Fällen, in denen prozessgenerierte Daten verwendet werden, handelt es sich in der Regel um Internetprotokolle, also um einen bestimmten Typus prozessgenerierter Daten. Unternehmensdatenbanken werden dagegen beispielsweise praktisch nie ausgewertet. Ebenso selten konstruieren Forscher nachträglich Datensätze.

Ihrer Natur nach erfasst die standardisierte Befragung bewusste Einstellungen und Meinungen von Personen. Unbewusste Einstellungen und Dinge, die der Befragte nicht verbalisieren kann oder nicht zugeben mag,

<sup>11</sup> Hinzu kommt ein weiteres Problem: Es ist gängige Praxis, das Signifikanzniveau auf maximal  $\alpha = 0,05$  anzusetzen. Das bedeutet aber, dass 5 % der Stichproben fehlerhafte Ergebnisse liefern. Bei der Menge der quantitativen Datensätze, die mittlerweile in den Sozialwissenschaften erstellt werden, muss man wahrscheinlichkeitstheoretisch davon ausgehen, dass eine große Zahl von Datensätze fehlerhaft sind.

<sup>12</sup> Die schließende Statistik sagt nichts über die empirische Relevanz der Ergebnisse aus – das können nur die deskriptive Statistik, qualitative und historische Sozialforschung leisten. Ob empirisch starke Zusammenhänge auch theoretisch relevant sind, hängt vom Erkenntnisinteresse ab. Hierzu muss der Forscher die Daten interpretieren. Dies ist also keine statistische sondern eine soziologische Frage (*Baur (2003c)*).

können über diese Erhebungsform nicht erfasst werden. Ebenso wenig sagen Einstellungen und Meinungen etwas darüber aus, wie sich die Befragten real verhalten.<sup>13</sup>

### b) *Fixierung auf Individualdaten*

Mit der Beschränkung auf die standardisierte Befragung geht eine weitere Selbstbeschränkung der meisten empirisch arbeitenden Soziologen (im Gegensatz zu Politikwissenschaftlern und Ökonomen) einher: die Fixierung auf Menschen als Individuen. Eine der wenigen Ausnahmen in der Soziologie ist die bereits oben erwähnte Unternehmensdatenbank von *Windolf*.

Die Beschränkung auf Umfragedaten ist problematisch: Kollektive Phänomene entstehen oft aus unbewussten Nebenfolgen des Handelns. Entsprechend sagen Umfragedaten nur begrenzt etwas über kollektive Phänomene aus. Benötigt werden deshalb auch Daten über soziale Kollektive, wie Unternehmen, Länder, Gemeinden, Situationen, Texte usw. Diese werden nur selten erhoben und beschränken sich dann meist auf allgemeine Indikatoren, wie Preisindizes, Insolvenzrate, BSP usw.<sup>14</sup> Veränderungen von Handlungsmustern und Veränderungen auf der Kollektivebene bleiben ausgeklammert (*Schulze* (2001a), *Lieberson* (1998): 134). Vielen Forschern ist gar nicht bewusst, dass sie mit ihren Daten und Verfahren gar nicht Schlüsse über die Analyseebene ziehen können, die sie anstreben. Ohne es zu merken, ändern sie im Lauf der Forschung ihr Analyseziel, z. B. wenn sie interkulturelle Unterschiede analysieren wollen, analysieren sie schließlich interindividuelle Unterschiede. Viele der Variablen, die sie interessieren, können weder direkt beobachtet noch gemessen werden (*Mayer* (1998)).

Diese Fokussierung auf die Individualebene ist umso fataler, als sich ein großer Teil der soziologischen Theoretiker für makrosoziale Phänomene interessiert. Für diese Theoretiker wären gerade die ausgeklammerten Bereiche der empirischen Sozialforschung die interessantesten. Methoden empfinden sie entsprechend mehr als Einschränkung als als Hilfe für ihre eigenen Erkenntnisinteressen. Dies wiederum ist einer der Gründe dafür, dass sich viele theoretisch interessierte Sozialwissenschaftler von den Methoden abwenden oder sie gar nicht erst erlernen.<sup>15</sup>

### c) *Fixierung auf bestimmte Auswertungsverfahren*

Um wirklich das volle Informationspotential ihrer Daten auszuschöpfen, müssten Forscher ihre Auswertungsverfahren variieren: Sie müssten jeweils das dem spezifischen Datensatz und der Forschungsfrage angemessene

<sup>13</sup> Ein Beispiel ist die Bundestagswahl 2002: In Meinungsumfragen signalisierten die Wahlberechtigten klar, dass sie die Schwarz-Gelb bevorzugten. Bei der Bundestagswahl selbst wählte die Mehrheit Rot-Grün. Haben die Befragten kurzfristig ihre Meinung geändert? Haben sie einfach in den Umfragen gelogen, um die Regierungskoalition zu warnen? Oder ist es für die Befragten einfach etwas anderes, in einer Umfrage eine Meinung zu äußern, als auf dem Wahlformular sein Kreuz zu machen? Jedenfalls haben Viele nicht die Partei gewählt, von der sie gesagt haben, dass sie sie wählen würden.

<sup>14</sup> Um es noch mal zu betonen: Sowohl Umfragedaten als auch allgemeine Indikatoren über Kollektive haben ihren Nutzen. Es wäre unsinnig, auf ihr Erkenntnispotential zu verzichten. Die Soziologie beschränkt sich aber ihrem Erkenntnispotential, wenn sie auf Beobachtungsdaten, nachträglich konstruierte Daten und Daten über Kollektive verzichtet.

<sup>15</sup> Beispielsweise versteht *Talcott Parsons* unter empirischer Forschung nur quantitative Umfrageforschung sowie die Datenerhebungstechniken Beobachtung und nicht standardisiertes Interview. Letztere betrachtet er aber nicht als soziologische sondern als anthropologische Methoden (*Parsons* (1975a): 47). Für eine Studie über den Arztberuf in den 1930ern entnahm er sein empirisches Material hauptsächlich über teilnehmende Beobachtung und nicht standardisierte Interviews (*Parsons* (1975a): 12, 47). Ende der 1950er führte er zusammen mit *Florence Kluckhohn* und *Linie Stouffer* eine standardisierte Umfrage über die Mobilität von Highschool-Schülern durch. Die Ergebnisse beider Projekte wurden nie veröffentlicht. In den 1960ern startete er zusammen mit *Gerald Platt* ein zweites Umfrageprojekt – dieses Mal über die Akademiker in den USA (*Parsons* (1975a): 46-48). Zusätzlich arbeitete *Parsons* auch in seinem übrigen Werk empirisch. Offensichtlich betreibt er insbesondere die Verfahren der Metaanalyse verschiedener Theoretiker sowie seine historisch-vergleichenden Studien (*Parsons* (1972), *Parsons* (1975b)). *Parsons* empfand den Zwang, empirische Sozialforschung in diesem engen Verständnis zu betreiben, immer als Druck. So schreibt er in seiner Autobiographie: „Es dürfte verständlich sein, daß ein Sozialwissenschaftler wie ich, der sich so sehr in Dinge von allgemein theoretischer Bedeutung vertiefte, unter einem gewissen Druck steht aufgrund des Umstandes, daß man in Amerika mehr von solider empirischer Forschung hält“ (*Parsons* (1975a): 47). Er konnte mit den Forschungsmethoden, wie sie allgemein gelehrt wurden, nicht viel anfangen, weil er keine Verbindung zwischen seinen theoretischen Interessen und empirischer Sozialforschung sah. Auf die Möglichkeit einer solchen Verbindung hoffte er erst spät: „Vor allem seit ich *Stouffer* näher gekommen war [also ab den 1950ern], stieg mein intellektueller Respekt vor der empirischen Sozialforschung mächtig an, und damit meine Hoffnung, daß noch engere Verbindungen zwischen diesen Techniken [der empirischen Sozialforschung] und der Art Theorie, die mir am Herzen lag, ausgearbeitet werden können.“ (*Parsons* (1975a): 47).

Ähnlich wie *Parsons* hatte *Norbert Elias* seine Probleme mit der empirischen Sozialforschung. Obwohl *Elias* in seine gesamte Theorie mit Hilfe von historischen Analysen untermauert, empfand er dies nicht als empirische Methode. Als er in den 1960ern einmal mit quantitativen Daten hantierte, meinte er selbstironisch, dass er ja jetzt unter die Empiriker gegangen sein (persönliche Notiz von *Herman Korte*).

sene Verfahren anwenden. Häufig müssten sie mehrere multivariate Verfahren mit Verfahren der zeitlichen Analyse kombinieren. Die meisten Forscher spezialisieren sich dagegen auf ein einziges Verfahren. Dies ist in der Regel das für die Forschungstradition typische Verfahren.

Um nur einige Beispiele zu nennen: Die Bevölkerungswissenschaft und die Lebenslauf- und Berufsforschung konzentrieren sich auf einzelne Menschen. Die Methoden der Wahl sind die Ereignis- und Kohortenanalyse. Die Wirtschaftswissenschaften konzentrieren sich auf Veränderungen auf der Makroebene und bevorzugen die Zeitreihenanalyse. Politikwissenschaftler treiben sehr häufig Mehrebenenanalysen mit einem Fokus auf der Kollektivebene. Organisationssoziologen konzentrieren sich auf die Beziehungen zwischen Menschen in Organisationen oder zwischen Organisationen untereinander. Sie betreiben Netzwerkanalyse. Zeitbezüge sind selten. Quantitative Zeitvergleiche sind in der deutschsprachigen Forschung selten. In anderen speziellen Soziologien bevorzugen Forscher qualitative Verfahren. Gemeinsam ist diesen Teilgebieten der Menschenwissenschaften, dass irgendwann bestimmte Methoden entwickelt wurden, um konkrete Probleme zu lösen. Im Lauf der Zeit werden diese Verfahren zum Selbstzweck: Andere Verfahren und Blickwinkel werden zur absoluten Ausnahme.

### 3 Grenzen der bisherigen Auswertungsverfahren

Die bislang genannten Probleme lassen sich relativ leicht beheben, indem die Lehr- und Forschungspraxis geändert wird. Auch der folgende Problemkomplex lässt sich leicht beheben. Hierzu ist allerdings die Weiterentwicklung der bisherigen Auswertungsverfahren für quantitative Längsschnittsanalysen notwendig. Die quantitative Sozialforschung hat zwar im Bereich der Auswertungsverfahren in den vergangenen zwanzig Jahren enorme Fortschritte gemacht. Nach wie vor sind die verfügbaren Auswertungsverfahren zu begrenzt. Insbesondere in folgenden Bereichen sehe ich noch Lücken:

- 3) *Identifikation von Stabilitäten*: Die meisten Auswertungsverfahren identifizieren hervorragend Wandel, erkennen aber Stabilitäten schlecht (Mayer (1998)).
- 4) *Dynamisierung von Mehrebenenanalysen*: Um das Wechselspiel zwischen individuellen und kollektiven Handlungsmustern analysieren zu können, müssten die Verfahren der Mehrebenenanalyse (Engel (1998)) dynamisiert werden.
- 5) *Verschränkung von Zeitschichten*: Verschiedene soziale Phänomene wandeln sich unterschiedlich schnell – mehrere Zeitschichten (Koselleck (2000)) überlagern sich. Intuitiv mag man annehmen, dass Handlungsebene und Wandlungstempo korrespondieren. Dies ist aber nicht unbedingt der Fall. Hierzu ein Beispiel: Auf Finanzmärkten überlagern sich sehr kurzfristige Börsenschwankungen (Minuten- oder Stundentakt) mit mittelfristigen Konjunkturentwicklungen (ein oder mehrere Jahre) und sehr langfristigen Entwicklungen (Jahrzehnte). Bis auf die Zeitreihenanalyse erfasst kein statistisches Verfahren unterschiedliche Wandlungstempi gut.
- 6) *Verstärkte Analyse von Wandlungsmustern*: Die Sequenzanalyse und Optimal Matching-Analyse erfassen Wandlungsmuster sehr gut (Abbott (1995)). Diese Verfahren werden in den Sozialwissenschaften noch sehr selten angewandt und sind entsprechend nicht besonders weit entwickelt. Hierbei können die Sozialwissenschaften von den von Informatikern entwickelten Data Mining-Techniken lernen (Fromm (2003), Srivastava (2004)).
- 7) *Dynamisierung multivariater Verfahren*: Eine Reihe von multivariater Verfahren kann komplexe Muster auf der Querschnittsebene entdecken. Beispielsweise eignet sich die Clusteranalyse zur Identifikation von Typen, die Faktorenanalyse zur Skalenbildung usw. Der nächste Entwicklungsschritt sollte sein, diese Verfahren zu dynamisieren.

## 4 Begrenzung durch die Datengrundlage

Die bisher angesprochenen Probleme quantitativer Längsschnittsanalysen lassen sich relativ leicht beheben – entweder durch Verbesserung der Forschungspraxis oder durch Weiterentwicklung des methodischen Arsenals. Das Problem, auf das ich nun eingehe, kann unter Umständen eine erste harte Grenze quantitativer Längsschnittsanalysen sein. Es handelt sich hierbei um die Begrenzung durch die Datengrundlage: Es kann sein, dass schlichtweg keine Daten existieren, um eine *bestimmte* Forschungsfrage zu beantworten oder ein spezifisches Verfahren anzuwenden. Mit anderen Worten: Die Begrenzung der Datengrundlage muss nicht, kann aber für einzelne Projekte ein Problem darstellen. Was bedeutet „Begrenzung der Datengrundlage“ nun im Einzelnen, und was kann man dagegen tun? Grob lassen sich zwei Kategorien unterscheiden: Stichprobenprobleme, die speziell bei quantitativen Längsschnittsuntersuchen auftauchen, sowie ein Mangel an Daten überhaupt.

### 4.1 Spezielle Stichprobenprobleme bei quantitativen Längsschnittsanalysen

Da sich für Längsschnittsanalysen geeignete Datensätze in der Regel aus einer Folge von Querschnittsdatsätzen zusammensetzen, multiplizieren sich etwaige Stichprobenprobleme dieser Querschnittsdatsätze. Hinzu kommen einige spezielle Stichprobenprobleme, die nur bei Längsschnittsanalysen auftauchen. Diese sind die Vergrößerung des Konfidenzintervalls, Lücken in der Protokollierung und die Veränderung der Grundgesamtheit.

#### a) Vergrößerung des Konfidenzintervalls

Eine Reihe von Stichprobenproblemen gelten nur für ein spezifisches Erhebungsdesign, lassen sich also beheben, indem man auf ein anderes Erhebungsdesign zurückgreift (vgl. *Tabelle 1*). Beim Trenddesign und bei Replikationsstudien ist beispielsweise die Wahrscheinlichkeit besonders groß, dass scheinbare Veränderungen tatsächlich Messfehlern sind. Das Stichprobenproblem stellt sich doppelt: Der Forscher muss fragen, ob jede der einzelnen Stichproben verallgemeinerbar sind und ob die Stichproben vergleichbar sind. Zufällige Schwankungen können dazu führen, dass die Stichprobenfehler zu zwei Zeitpunkten gegenläufig sind. Hierzu ein hypothetisches Beispiel aus der Wahlforschung: Zu Zeitpunkt t1 geben 40 % und zum Zeitpunkt t2 nur noch 30 % der Befragten an, dass sie SPD wählen würden. Es ist nun möglich, dass die SPD tatsächlich massiv an Zustimmung verloren hat. Es kann aber auch sein, dass zufällig das erste Mal hauptsächlich CDU-Anhänger, das zweite Mal hauptsächlich SPD-Anhänger befragt wurden. Selbst wenn in beiden Fällen Zufallsstichproben vorliegen, vergrößert sich so das Konfidenzintervall bei Schätzungen.

#### b) Lücken in der Protokollierung

Bei allen Erhebungsdesigns existieren Lücken in der Protokollierung – allerdings in jeweils unterschiedlicher Form (vgl. *Tabelle 1*). Diese Lücken führen zu systematischen Verzerrungen der Forschungsergebnisse. Wie groß der Bias ist, muss im Einzelfall abgeschätzt werden. Hierzu muss man die für das spezifische Erhebungsdesign typischen Protokollierungsprobleme sowie die Umstände der jeweiligen Untersuchung kennen. Protokollierungsprobleme entstehen bei der Datenproduktion und Datenaufbewahrung.

Wie bereits oben erwähnt, ist auch die Erhebung quantitativer Querschnittsdaten lückenhaft. Diese Fehler multiplizieren sich bei Längsschnittsanalysen. Welcher Art sie sind, hängt vom jeweiligen Designtyp ab. Bei der Retrospektivbefragung, bei replizierbaren Querschnittsdaten sowie beim Panel- und Trenddesigns entstehen systematische Fehler durch den Auswahlprozess, durch Ausfälle und durch fehlende Werte. Aber auch bei prozessgenerierten qualitativen und quantitativen Daten sind die Aufzeichnungen in der Regel lücken- und fehlerhaft. Welcher Art diese Fehler sind, hängt vom Datentyp ab. Beispielsweise erhalten Zeitungsdaten einen guten Teil des Tagesgeschehens – nur eben nicht auf systematische Weise. Sie berichten darüber, was Journalisten zum jeweiligen Zeitpunkt wussten und für wichtig hielten. Nur, weil über etwas nicht berichtet wird, heißt das nicht, dass es nicht geschehen ist. Anders stellt sich die Lage beispielsweise bei Formularen z. B. des Einwohnermeldeamts dar. In verschiedenen Städten können die Formularfelder unterschiedlich sein. Dies umso wahrscheinlicher, je weniger standardisiert der Protokollierungsprozess ist. Doch selbst gleiche Felder sichern nicht vor Protokollierungslücken: Jeder, der einmal für einen Kongress oder eine Lehrveranstaltung Anmeldeformulare ausfüllen hat lassen, weiß, wie viele Möglichkeiten existieren, ein scheinbar simples Formu-

lar auszufüllen. Nicht nur das – immer wieder vergessen einzelne Personen, bestimmte Formularfelder auszufüllen.

**Tabelle 1: Stichprobenprobleme und Erhebungsdesign**

Erhebungsdesign	Zeitfenster für generelle Verfügbarkeit	Vergrößerung des Konfidenzintervalls	Protokollierungslücken		
			Datenproduktion (Querschnitt)	Datenproduktion (Längsschnitt)	Datenselektion
<i>Retrospektivbefragung</i>	Vergangene 70 Jahre (bzw. etwa menschliche Lebensspanne)				Keine oder falsche Erinnerung
<i>Paneldesign</i>	Seit Mitte der 1980er		<ul style="list-style-type: none"> <li>• Selektivität des Auswahlprozesses</li> <li>• Ausfälle</li> <li>• Fehlende Werte</li> </ul>	<ul style="list-style-type: none"> <li>• Panelmortalität</li> <li>• Spell-Effekte</li> </ul>	Kontextabhängig, Multiplizierung mit der Zeit:
<i>Trenddesign</i>	Seit Mitte der 1970er	X		[individueller Wandel kann nicht nachvollzogen werden]	
<i>Replizierbare Querschnittsdaten</i>	Seit 1940er	X			
<i>Prozessgenerierte quantitative Daten</i>	<ul style="list-style-type: none"> <li>• Verwaltungsdaten / wirtschaftliche Maßzahlen: seit 19 Jh.</li> <li>• Verwaltungs- und Firmendatenbanken: seit Mitte 1960er</li> <li>• Internetprotokolle: seit Anfang 1990er</li> </ul>		Abhängig vom Datentyp, z. B. <ul style="list-style-type: none"> <li>• bei Akten: unterschiedliche oder nicht ausgefüllte Formularfelder</li> <li>• bei Zeitungsdaten: Journalisten erachten eine Vorfälle für nicht wichtig</li> </ul>	Abhängig vom Datentyp, z. B. <ul style="list-style-type: none"> <li>• bei Internetprotokollen: Serverausfälle</li> <li>• bei Zeitungsdaten: wandelnde Bedeutung eines Themas in der Berichterstattung</li> </ul>	<ul style="list-style-type: none"> <li>• Natürlicher Zerfall, weil kein Interesse an Daten besteht</li> <li>• Bewusste Zerstörung</li> </ul>
<i>Nachträglich konstruierte Datensätze</i>	Immer, sofern qualitativ auswertbares Material vorliegt				

Erstreckt sich die Datenproduktion über einen längeren Zeitraum (was bei Erhebungsdesigns für Längsschnitte die Regel ist), entstehen zusätzliche Fehler. Auch diese sind von Erhebungsdesign zu Erhebungsdesign unterschiedlich. Typische Probleme bei der Panelanalyse sind die Panelmortalität<sup>16</sup> und Spell-Effekte.<sup>17</sup> Bei der laufenden Protokollierung können solche Lücken entstehen, indem zwischendurch vergessen wurde, Informationen aufzuzeichnen. Gründe hierfür könnten bei Akten und Datenbanken ein Personalwechsel sein oder dass jemand krank war. Bei Internetprotokollen kann schlicht der Server ausfallen oder das Programm abstürzen. Bei Zeitungen kann sich unter anderem die Art der Berichterstattung ändern. Beispielsweise achten deutsche Zeitungen seit ein paar Jahren wieder verstärkt auf Korruptionsfälle in der öffentlichen Verwaltung und in Füh-

<sup>16</sup> Von Panelwelle zu Panelwelle fällt in der Regel ein Teil der Individuen aus. Bei Personen sind typische Gründe z. B. gesunkene Teilnahmemotivation, Krankheit, Tod und Adressänderung durch Umzug. Die Panelmortalität kann in natürliche und nicht natürliche Faktoren geteilt werden. Unter natürlichen Faktoren versteht man den Tod oder den Wohnungswechsel von Befragten, d.h. die Abwesenheit der Personen (noncoverage error). Die nicht natürlichen Faktoren betreffen die Verweigerung der Befragung (nonresponse error). Während die natürliche Panelmortalität kaum Verzerrungen verursacht, da sie die normale Bevölkerungsentwicklung widerspiegelt, schafft die nicht-natürliche größere Probleme für die Repräsentativität. Auch neue Teilnehmer in das Panel aufzunehmen, löst das Problem nicht – für diese Personen fehlen Messungen zu früheren Zeitpunkten.

<sup>17</sup> Dadurch, dass nur zu bestimmten Zeitpunkten gemessen wird, entstehen zwei Typen von blinde Flecken: (1) Es kann sein, dass zwischen den Messzeitpunkten eine Veränderung eintrat, aber der ursprüngliche Zustand wieder hergestellt wurde. Beispielsweise wurde jemand arbeitslos, hat aber mittlerweile wieder eine Beschäftigung gefunden. Wenn nicht gezielt nach solchen Veränderungen gefragt wird, gewinnt man durch diese Datentypen den Eindruck, dass seit dem ersten Erhebungszeitpunkt alles gleich geblieben ist (= Short-Spell-Problem). Das zweite Problem dieser Erhebungsdesigns ist, dass – selbst sich etwas verändert hat und man dies auch feststellen kann – man nicht weiß, wann die Veränderung eingetreten ist. Wie lange dauerte also der frühere Zustand an, wann begann der neue (*Steinhage (2000)*)? Hierzu ein Beispiel: In einer Panel-Befragung werden die Befragten alle vier Jahre befragt. Person A hatte bei der ersten Befragung eine Stelle, ist aber jetzt arbeitslos. Wie lange ist er arbeitslos? Seit vier Jahren? Dann wäre er ein Langzeitarbeitsloser. Oder erst seit einer Woche? Dann befindet er sich möglicherweise nur zwischen zwei Stellenwechseln. Um dieses Problem zu umgehen, benötigt man Ereignisdaten. Diese entstehen durch laufende Protokollierung oder durch Retrospektivbefragung (*Blossfeld / Hamerle / Mayer (1986)*).

rungetagen von Firmen. Dass in den Jahren davor nicht über sie berichtet wurde, bedeutet nicht, dass sie nicht existierten, sondern nur, dass Journalisten sich hierfür nicht interessierten.<sup>18</sup>

Bei Längsschnittanalysen taucht schließlich eine Fehlerquelle auf, die bei Querschnittanalysen meist nicht auftaucht, weil Datenerhebung und -auswertung kurz aufeinanderfolgen. Für Längsschnittanalysen müssen dagegen Daten über Jahre, Jahrzehnte oder gar Jahrhunderte aufbewahrt werden. Je mehr Zeit vergeht, desto größer ist die Gefahr, dass Daten verloren gehen. Auch dieser Prozess verläuft nicht zufällig: Einerseits können Menschen Daten bewusst zerstören – und in der Regel sind dies bestimmte Daten. Man denke nur an die verschwundenen Akten aus dem Kanzleramt. Andererseits unterliegen Datenträger dem natürlichen Zerfall. Bücher verschimmeln, Disketten werden fehlerhaft usw. Menschen müssen sich anstrengen, um diesen Zerfall entgegenzuwirken. Dies tun sie nur, wenn Ihnen Informationen wichtig sind. Was bewusst zerstört und was aufbewahrt wird, hängt vom jeweiligen soziokulturellen und politischen Kontext ab. Da dieser sich wandelt, ändert sich auch das, was zerstört und aufbewahrt wird. Daraus folgt zweierlei: Erstens muss man den jeweiligen Kontext kennen, um die Art des Selektionsbias abschätzen zu können. Zweitens wird die Datengrundlage immer schlechter, je weiter man die Vergangenheit zurückgeht. Irgendwann kommt der Punkt, an dem Forscher dankbar sein dürfen, überhaupt Daten zu finden – egal welcher Art.

Retrospektivbefragungen sind keine Lösung für dieses Problem, und zwar aus zweierlei Gründen: Erstens ist die menschliche Lebensspanne begrenzt. Will man etwas über Ereignisse erfahren, die weiter in der Vergangenheit liegen, lebt niemand mehr, den man befragen könnte. Zweitens ist – wie Historiker immer wieder betonen – das menschliche Gedächtnis eine der unzuverlässigsten Datenträger überhaupt: Jeder Mensch nimmt seine Umgebung bereits selektiv wahr, bekommt bestimmte Ereignisse also erst gar nicht mit. Menschen deuten Ereignisse unterschiedlich, verknüpfen sie auf unterschiedliche Weise miteinander – was für den einen eine angeregte Unterhaltung war, war für den anderen ein Streit. Menschen vergessen Ereignisse. Selektion, Interpretation und Erinnerung sind zudem unscharf und von Mensch zu Mensch verschieden: Menschen haben unterschiedlich gute Wahrnehmungs-, Deutungs- und Erinnerungsfähigkeiten. Ob wir Geschehnisse im Gedächtnis bewahren, hängt zudem davon ab, wie wichtig sie uns waren: Zwar erinnern Menschen sich in der Regel an für sie bedeutsame Ereignisse noch nach Jahrzehnten genau, andere Ereignisse vergessen sie aber innerhalb von Momenten. Was wir wahrnehmen, wie wir es interpretieren und ob wir uns daran erinnern (wollen), hängt von unserer Perspektive ab. Diese bestimmt, was wir für wichtig erachten – unwichtige Informationen werden schnell ausgewählt. Hinzu kommt, dass wir unsere Perspektive wechseln können. Mit einem Perspektivenwechsel werden auch vergangene Ereignisse neu interpretiert. Ein typisches Beispiel hierfür sind zerbrochene Partnerschaften. Im Nachhinein deuten Ex-Partner häufig Vieles, was sie früher nicht so schlimm fanden, als schon immer vorhandenes, gravierendes Problem um. Schöne Erlebnisse blenden sie dagegen aus. Oft werden neue Ereignisse hinzugedichtet, die so nie geschehen sind. Stets muss sich der Forscher also fragen: Hat der Befragte etwas nicht erlebt? Erinnert er sich einfach nicht? Erinnert er sich richtig (*Duby* (1988): 8)? Bereits der einzelne Mensch ist also eine höchst unzuverlässige Datenquelle – das ist ja der Grund, warum Menschen überhaupt begonnen haben, Ereignisse schriftlich zu fixieren. Dies gilt umso mehr, wenn sich der Forscher – wie in den Sozialwissenschaften – für Gruppen von Menschen interessiert. Ob Meinungen und Erinnerungen von Menschen vergleichbar sind, muss von Fall zu Fall entschieden werden, ist aber immer höchst fraglich.

### c) Wandel der Grundgesamtheit

Unabhängig vom Erhebungsdesign existiert ein Stichprobenproblem, das kein Problem der Stichprobenziehung sondern ein Problem der *Grundgesamtheit* ist. Diese kann sich nämlich ebenfalls wandeln. Oft geschieht dies unmerklich: Menschen sterben, andere werden geboren und altern. Unternehmen werden gegründet und gehen in Konkurs. Die Grundgesamtheit kann sich aber auch sehr plötzlich wandeln. Das prominenteste Beispiel ist die Wiedervereinigung: Innerhalb eines Jahres haben sich das geographische Gebiet und die Bevölkerung Deutschlands deutlich vermehrt. Die stellt ein methodisches Problem dar: Kann man Umfragen aus der Zeit vor der Wiedervereinigung mit Umfragen nach der Wiedervereinigung vergleichen? Ist es legitim anzunehmen, dass die Westdeutschen vor 1989 den Westdeutschen nach 1989 entsprechen? Keine dieser Fragen kann einfach mit „Ja“ beantwortet werden. Die Ostdeutschen haben ja nicht dieselben Lebenserfahrungen gemacht wie die Westdeutschen, können also nicht einfach mit ihnen in einen Topf geworfen werden. Umgekehrt hat die Wiedervereinigung auch die Westdeutschen verändert. Hinzu kommt, dass seitdem die Binnenwanderung (insbesondere von Ost- nach Westdeutschland) stark zugenommen hat, so dass heute viele Westdeutsche in Ostdeutschland und noch mehr Ostdeutsche in Westdeutschland leben.

Dieses Beispiel verdeutlicht, dass es erforderlich ist, sich bei jeder Längsschnittanalysen drei grundsätzliche Fragen zu stellen: Hat sich die Grundgesamtheit gewandelt? Wenn ja, was bedeutet dies für das Erkenntnisinteresse? Und welche Stichprobe ist repräsentativ für die Grundgesamtheit?

<sup>18</sup> Dass das Trenddesign in dieser Liste nicht genannt wird, ist kein Vorteil dieses Erhebungsdesigns sondern ein Nachteil: Mit dem Trenddesign ist es an sich nicht möglich, individuellen Wandel zu verfolgen. Bei einigen der anderen Designtypen ist dies dagegen prinzipiell möglich, auch wenn die Analyse dieses Wandels lückenhaft und verzerrt sein kann.



Dieses Problem lässt sich weder grundsätzlich noch abstrakt sondern nur am Einzelfall lösen. Eine erste Möglichkeit, mit ihm umzugehen, ist, alle Analysen dreifach zu machen: jeweils einmal mit der alten, einmal mit der neuen und einmal von der alten auf die neue Grundgesamtheit. Hat man Glück, sind die Ergebnisse stabil – der Wandel der Grundgesamtheit stellt kein Interpretationsproblem dar. Wenn sich die Ergebnisse unterscheiden, bleiben nur qualitative Verfahren, um die Ursache für diese Differenzen zu erforschen.<sup>19</sup>

## 4.2 Mangel an Daten

Stichprobenprobleme sind nicht die einzigen Probleme, vor denen Forscher stehen können, die quantitative Längsschnittsanalysen betreiben wollen. Es kann auch passieren, dass für die spezifische Forschungsfrage überhaupt keine Daten existieren. Dieser Datenmangel manifestiert sich in einer Reihe unterschiedlicher Formen.

### a) Wandel der Forschungspraxis

Zunächst kann der Datenmangel einfach daher rühren, dass sich die Forschungspraxis wandelt. Etwas eigentlich Positives – der Fortschritt der Methoden – stellt also für die Forschung ein Problem dar. Kritisch sind die Erfindung neuer Erhebungsdesigns, der Wandel der Messinstrumente und der Wandel der Auswertungsverfahren:

Banalerweise können Daten eines bestimmten Designtyps erst existieren, nachdem dieses Erhebungsdesign erfunden wurde. Durch den Fortschritt der empirischen Sozialforschung ergibt sich eine zeitliche Staffelung (vgl. *Tabelle 1*): Manche Indizes (wie die Bevölkerungsgröße und die Zahl der Arbeitsstätten und Berufe) werden seit Anfang des 18. Jahrhunderts im Rahmen der amtlichen Statistik erhoben. Umfragedaten im Querschnittsdesign liegen frühestens seit 1940ern vor. Seit etwa Mitte der 1960er halten Unternehmen und die öffentliche Verwaltung Informationen in Datenbanken fest. Seit etwa Mitte der 1970er wird das Trenddesign, seit etwa Mitte der 1980er das Paneldesign systematisch angewandt. Seit Mitte der 1990er kommen Internetprotokolle hinzu. Bei Retrospektivbefragungen ergibt sich eine andere Form der zeitlichen Reichweite: Befragt werden können nur Menschen, die noch leben. Geht man davon aus, dass es auch heute noch sehr schwer ist, Über-80-Jährige zu finden und dass sich diese an die ersten zehn Jahre ihres Lebens nur schwer erinnern, kann man mit Retrospektivbefragungen zeitlich maximal etwa 70 Jahre zurückgehen. Der einzige Designtyp, der potentiell keiner zeitlichen Beschränkung unterliegt, ist die nachträgliche Konstruktion von Datensätzen mit Hilfe qualitativer Verfahren. Praktisch kann jedoch auch dieser Designtyp begrenzt sein: Je weiter man in die Vergangenheit zurückgeht, desto schwerer ist es – wegen des bereits erwähnten Selektivitätsbias – überhaupt Daten irgendeiner Art zu bekommen. Durch die bereits erwähnte Fixierung der meisten Sozialwissenschaftler auf Individualdaten kann es insbesondere schwer sein, Veränderungen von kollektiven Handlungsmustern nachzuzeichnen. Schließt man von individuellen auf kollektive Merkmale, ergibt sich die Gefahr des individualistischen Fehlschlusses (*Schulze (2004)*).

Nicht nur im Bereich der Erhebungsdesigns, auch im Bereich der Messinstrumente kommt es zu Methodenfortschritten. Nehmen wir z. B. die geschlossene Befragung: Es kann sein, dass man im Lauf der Zeit feststellt, dass eine frühere Frageformulierung schlecht war. Dies stellt ein Dilemma dar: Behält man die Frage bei, produziert man weiter wertlose Daten. Ändert man sie, fängt man wieder bei Null an: Man kann nicht davon ausgehen, dass sich unterschiedliche Formulierungen im Fragebogen vergleichen lassen. Die neue Formulierung ist also als eine neue Querschnittserhebung zu betrachten.

Schließlich wurden in den vergangenen Jahren neue, leistungsfähige Auswertungsverfahren entwickelt, und es wird wohl auch künftig zu Fortschritten in diesem Bereich kommen. Häufig erfordern aber neue Auswertungsverfahren einen bestimmten Fragetyp. Bei Trenddesigns und Replikationsstudien ist es beispielsweise unmöglich, individuelle Veränderungen festzustellen. Dadurch ergibt sich die Gefahr des ökologischen Fehlschlusses (*Schulze (2004)*). Ereignisanalysen kann man nur durchführen, wenn Ereignisdaten vorliegen. Daraus ergibt sich ein ähnliches Dilemma wie beim Wandel der Messinstrumente: Um den neuen Erkenntnissen gerecht zu werden, sollte man die Frageformulierung oder das gesamte Erhebungsdesign ändern. Damit werden aber frühere Erhebungen zum selben Thema für Zeitvergleiche wertlos.

### b) Unterschiedliche Zeitfenster

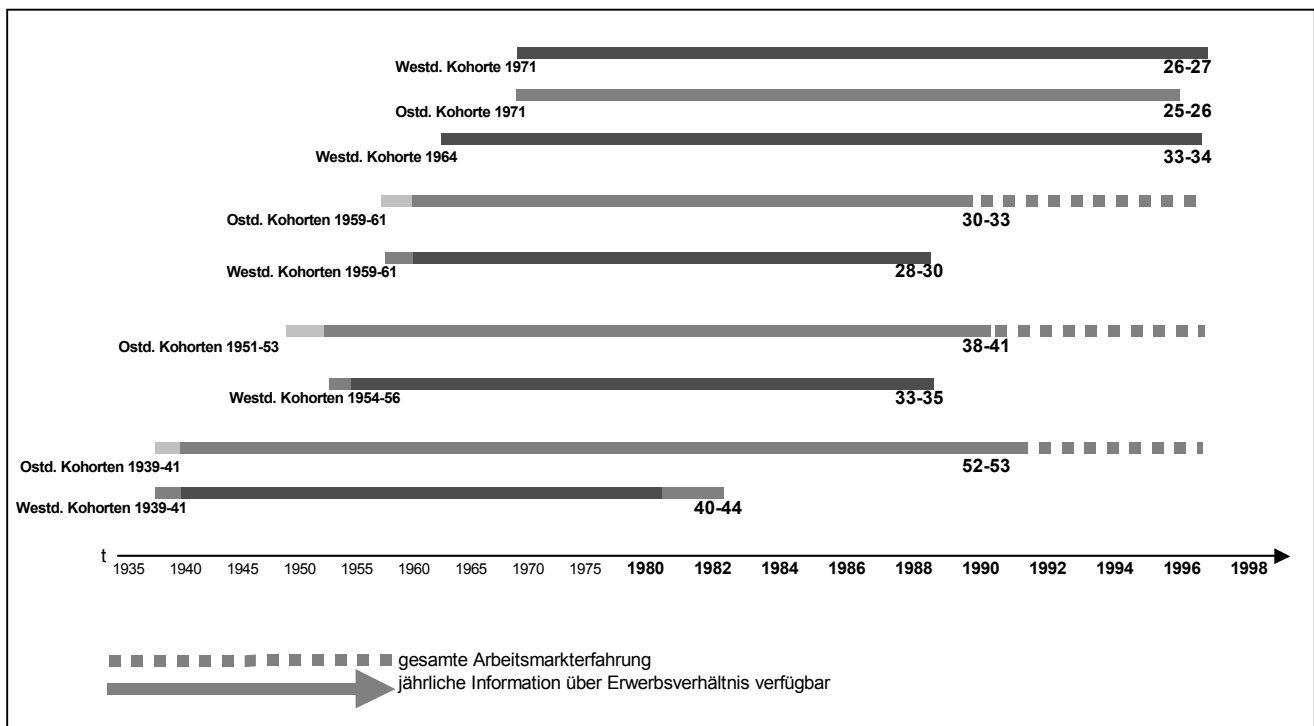
Häufig existieren zwar Daten zu einem bestimmten Forschungsthema, die erfassen aber unterschiedliche Zeitfenster. Will man beispielsweise die Erwerbskarrieren von west- und ostdeutschen Männern mit Hilfe der German Life History Study (GLHS) untersuchen, ergibt sich dabei das Problem, dass für verschiedene Geburtsko-

<sup>19</sup> Dieser Lösungsvorschlag stammt von *Karin Kurz*, der ich an dieser Stelle herzlich für den Hinweis danken möchte.

herten nur Daten für unterschiedliche Zeiträume existieren (*Grunow (2004)*; vgl. auch *Grafik 4*). Dies muss nicht unbedingt ein Problem sein: Will man beispielsweise die Erwerbskarrieren der Männer in mittleren Karrierephase untersuchen, stellt dies kein Problem dar: Diese Karrierephase findet bei unterschiedlichen Kohorten zeitlich versetzt statt, und für alle Zeitfenster existieren Daten. Will man dagegen analysieren, wie sich die Karrieren parallel (also im selben Zeitraum) in verschiedenen Altersgruppen entwickelten, kann man mit Hilfe dieser Daten nur die Zeit von Mitte der 1970er bis Anfang der 1980er untersuchen. Dieses Problem verschärft sich bei internationalen Vergleichen (*Grunow (2004)*).

Selbst wenn die Zeitfenster der Daten gleich sind, heißt dies noch nicht, dass die Daten auch vergleichbar sind: In unterschiedlichen Kollektiven oder für unterschiedliche soziale Gruppen kann das Timing sozialer Prozesse unterschiedlich sein. Beispielsweise weist *Grunow (2004)* darauf hin, dass in Dänemark zahlreiche Prozesse auf dem Arbeitsmarkt (wie z. B. die Öffnung gegenüber dem Weltmarkt und die Flexibilisierung der Beschäftigungsverhältnisse) bereits in den 1920ern abgeschlossen waren, die in Deutschland erst in den vergangenen zwanzig Jahren einsetzten. Je nachdem, worauf der Vergleich abzielt, kann es erforderlich sein, dieselben oder unterschiedliche Zeiträume zu vergleichen: Interessiert, ob die gegenwärtigen globalen Prozesse unterschiedliche Länder gleichermaßen beeinflussen, sollten sowohl für Dänemark als auch für Deutschland dieselben Zeitfenster (z. B. 1980 bis 2000) betrachtet werden. Interessiert dagegen, ob die Öffnung gegenüber dem Weltmarkt in beiden Ländern ähnliche Auswirkungen auf Erwerbskarrieren hatte, müssen unterschiedliche Zeitfenster analysiert werden (in diesem Fall 1980 bis 2000 in Deutschland mit 1910 bis 1920 in Dänemark). Geschichtswissenschaftler kennen dieses Problem. Will man beispielsweise die deutsche Industrielle Revolution mit der englischen, amerikanischen, russischen usw. vergleichen, sind sehr unterschiedliche Zeitfenster zu untersuchen, da verschiedene Länder zeitlich versetzt industrialisiert wurden. Alternativ kann man untersuchen, wie die Wirtschaftsentwicklung in anderen Ländern in der Zeit verlief, in der sich Deutschland industrialisierte.

**Grafik 4: Unterschiedliche Zeitfenster verfügbarer Daten am Beispiel des Kohortenvergleich West- und Ostdeutschland in der GLHS**



Quelle: *Grunow (2004)*, Grafik leicht modifiziert

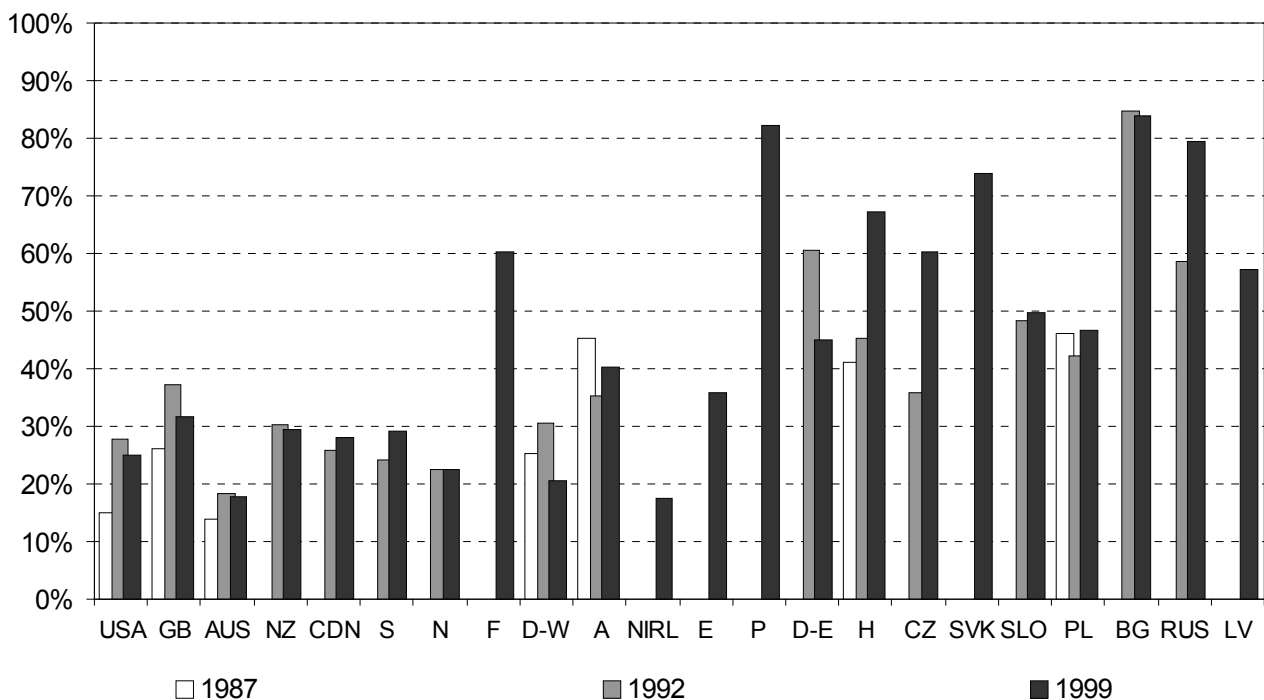
### c) Wandel des Kontextes

Weiter oben habe ich erläutert, dass bereits bei Querschnittsbefragungen der historisch-kulturelle Kontext zu beachten ist. Hierzu ein weiteres Beispiel: *Grafik 5* zeigt, welcher Anteil der Befragten in verschiedenen Ländern angibt, dass die soziale Ungleichheit in ihrem Land ihrer Ansicht nach zu groß sei. Die Datenquelle sind die ISSP-Wellen von 1987, 1992, 1999. Wie *Lück (2004)* betont, müssen diese Daten vorsichtig interpretiert werden. Insbesondere muss der Kontext beachtet werden: Betrachtet man nur die Grafik, könnte man auf die Idee kommen, dass die Westdeutschen (D-W), die Schweden (S) und die Amerikaner (USA) ein ähnliches Maß an sozialer Ungleichheit wünschen. Dies anzunehmen, ist allerdings unsinnig – die schwedischen Einkommen sind wesentlich gleichmäßiger verteilt als die amerikanischen, und der schwedische Sozialstaat verteilt wesent-

lich stärker um als der amerikanische. Deutschland liegt jeweils in der Mitte. Die Befragten antworten jeweils aus diesem Kontext heraus. Hieraus zu schließen, die Deutschen und die Schweden wünschten sich eine Anpassung an das amerikanische System, ist eine Umkehrung der Tatsachen. Dennoch geschieht genau dies häufig in der politischen Alltagsrhetorik.

Das Problem verschärft sich, wenn man die Zeitachse vergleicht. Für die USA und für Westdeutschland ist beispielsweise zu beobachten, dass zwischen 1987 und 1992 der Anteil derjenigen leicht anstieg, die das Ausmaß sozialer Ungleichheit für zu groß hielten. Danach sank dieser Anteil wieder. Dies könnten Stichprobenfehler sein, da es sich beim ISSP um ein Trenddesign handelt. Wenn keine Stichprobenfehler vorliegen, kann man aus diesen Daten dennoch nicht schließen, dass die Deutschen und Amerikaner ihre Haltung zu sozialer Ungleichheit geändert haben. Ebenso ist es möglich, dass die Einstellung zu sozialer Ungleichheit gleich geblieben ist, sich aber der Kontext gewandelt hat: Wenn eine Person ein bestimmtes Ausmaß sozialer Ungleichheit für akzeptabel hält, sich die (wahrgenommene) soziale Ungleichheit aber ändert, würde die betreffende Person eine andere Antwort geben. Entsprechend können die Befragungsdaten nur im Zusammenhang mit ihrem Kontext analysiert werden (Lück (2004)).

**Grafik 5: Wandel des Kontexts am Beispiel der Ansicht, dass im eigenen Land die soziale Ungleichheit zu groß ist**



Datenquelle: ISSP 1987, 1992, 1999 (gewichtet)

Frageformulierung (im englischen Fragebogen): „Do you agree or disagree ... Differences in income in [country] are too large.“

Antwortmöglichkeiten: „strongly agree“ (1) – (2) – (3) – (4) – (5) „strongly disagree“.

Quelle: Lück (2004)

Der Kontextwandel erstreckt sich nicht nur auf die Inhalte von Daten, sondern auch auf die Erhebungsinstrumente. Mit anderen Worten: Selbst wenn die Erhebungsinstrumente gleich bleiben, kann es sein, dass sie aufgrund des Kontextwandels Unterschiedliches messen. So kann sich die semantische Bedeutung eines Wortes oder einer Formulierung in Fragebögen sich ändern. Ehemals neutral oder positiv besetzte Wörter können ins Negative umschlagen. Beispielsweise bezeichneten Amerikaner im letzten Jahrhundert Schwarze allgemein als „Negros“. Der Gebrauch dieses Wortes gab keinerlei Hinweis auf die Einstellung der betreffenden Person gegenüber Schwarzen. Spätestens mit der amerikanischen Bürgerrechtsbewegung änderte sich dies: Der Gebrauch der Wörter „Negro“ und „Nigger“ wurde zum klaren Indiz für rassistische Einstellungen. Das neue neutrale Wort war „Black“. Bald bekam auch dieses Wort negative Konnotationen. Wer heute neutral über einen Amerikaner schwarzer Hautfarbe reden will, sollte tunlichst das Wort „African American“ benutzen. Dies gilt aber nicht für alle sozialen Räume und alle Bevölkerungskreise: Schwarze bezeichnen sich oft selbst noch als „Nigger“ oder „Negro“, allerdings mit selbstironischen oder aggressiven Untertönen. Sie benutzen das Wort auch allgemein als Schimpfwort gegenüber jeglichen Personengruppen. In Großbritannien ist das Wort „Black“ nach wie vor salonfähig.

Wörter und Formulierungen können nicht nur ihre Konnotation ändern. Ihre Bedeutung kann sich im Lauf der Zeit auch insgesamt ändern. Je abstrakter das Wort und je häufiger es benutzt wird, desto wahrscheinlicher ist dies. Ein Beispiel ist das Wort „Sozialstaat“. Zu einem bestimmten Zeitpunkt meinen verschiedene Personen damit Verschiedenes: Die einen verbinden damit heute alle staatlichen Leistungen, die nächsten nur die Kranken-, Renten- und Arbeitslosenversicherung, die nächsten nur die Sozialhilfe usw. Vergleicht man verschiedene Zeitpunkte, verschärft sich das Problem, denn der Sozialstaat war 1970 ein völlig anderer als heute.

Selbst scheinbar objektive Wörter können einen Bedeutungswandel erfahren. Seit kurzem können beispielsweise Homosexuelle heiraten – hat sich dadurch auch die Bedeutung des Wortes „Ehe“ verändert? Welche Bedeutung einem Wort genau zugeschrieben wird und ob sich diese Bedeutung ändert, müsste empirisch mit einer semantischen Analyse geprüft werden – also mit Hilfe eines qualitativen Verfahrens. Stellt man einen semantischen Wandel fest, stellt sich erneut die Frage: Darf man in einem solchen Fall die Formulierung im Fragebogen ändern? Muss man es nicht sogar?

Der Kontextwandel erstreckt sich noch auf einen anderen zentralen Bereich empirischer Sozialforschung: die Forschungsfrage. Weil sich Menschen wandeln, sind Forschungsinteressen und -themen selbst dem sozialen Wandel unterworfen (*Schulze (2002c)*). Dies wiederum kann zu Datenmängeln führen: Befand man früher Themen nicht für relevant, existieren heute auch keine Daten. Oder man würde heute andere Fragen stellen und andere Variablen miteinander in Verbindung setzen – die aber nicht erfasst sind. Beispielsweise werden die Transformation osteuropäischer Gesellschaften und die Globalisierungsfrage erst seit Beginn der 1990er in der diskutiert. Entsprechend kam vorher auch niemandem die Idee, zu diesen Themen Datensätze aufzubauen. Dasselbe gilt für zahlreiche Fragestellungen auf der Meso- und Makroebene. Oft interessieren jedoch andere Themen oder man möchte – gerade bei kultursoziologischen Untersuchungen – noch weiter in die Vergangenheit zurückgehen. Dies Problem besteht besonders bei komplexen Kausalzusammenhängen, bei Dimensionsvariablen und bei Typologien – also bei soziologisch relevanten Daten –, weil man hierzu viele verschiedene Variablen benötigt.

## 5 Theoretische Grenzen

Während sich die ersten drei Probleme quantitativer Längsschnittsanalysen größtenteils beheben lassen, ist das vierte quantitativen Längsschnittsanalysen immanent. Es handelt sich hierbei um grundsätzlich theoretische Grenzen. Diese liegen bei der Modellbildung für neue Forschungsfragen, komplexer Modellbildung sowie bei der Analyse von Deutungs- und Interaktionsmustern.

### 5.1 Neue Forschungsfragen

Die erste theoretische Grenze sind neue Forschungsfragen. Gerade in der Soziologie tauchen ständig neue Forschungsfragen auf, weil – wie ich bereits erwähnte – ihr Gegenstand dem sozialen Wandel unterworfen ist (*Schulze (2002c)*). Quantitative Verfahren eignen sich insbesondere für eine deduktiv-hypothesenprüfende Herangehensweise. Der Forscher muss also vor der Analyse zumindest wissen, welche Variablen relevant sind. Idealerweise sollten auch ein oder mehrere theoretisch ausgereifte Modelle über die Wirkungsweise der Variablen zur Verfügung stehen. Gerade bei neuen Themen, die noch relativ unerforscht sind, eignen sich in der Regel induktive, offene Herangehensweisen besser (*Schulze (2001b)*; *Schulze (2002b)*).

### 5.2 Komplexe Modelle

Zweitens können quantitative Verfahren nur angewandt werden, wenn sich die Modelle auf möglichst wenige Variablen reduzieren lassen. Oft hat man es in den Sozialwissenschaften aber mit komplexen Variablenzusammenhängen zu tun – und häufig sind gerade diese Interaktionen das, was interessiert.

### 5.3 Deutungs- und Interaktionsmuster

Bestimmte soziale Phänomene, die im Zentrum des Erkenntnisinteresses der Soziologie liegen, sind schwer erfragbar und kaum quantitativ zu fassen. Hier stoßen auch Verfahren der nachträglichen Konstruktion von Datensätzen an ihre Grenzen. Hierzu gehören Deutungsmuster und Interaktionsmuster – Verlaufsschemata, Episoden, Drehbücher, Zyklen, Muster, oder welchen Ausdruck Sie sonst verwenden wollen, kurz: typische Regelmäßigkeiten im sozialen Handeln, die sich nur in der Zeit manifestieren. Theoretisch kann man qualitativ Daten zwar so aufbereiten, dass sie kodiert und anschließend quantitativ analysiert werden können. Ein Beispiel hierfür ist die bereits erwähnte Rekonstruktion des Diskurses über die Arbeitslosigkeit in mehreren Bamberger Projekten. An diesem Beispiel zeigen sich aber drei grundsätzliche Probleme: Erstens liegt die Hauptarbeit in der qualitativen Aufarbeitung dieser Prozessmuster, bevor überhaupt ein Kodierschema entworfen werden kann. Zweitens greifen die quantitativen Daten immer zu kurz. Sie eignen sich lediglich, bereits identifizierte Muster zu überprüfen. Würde man sie alleine analysieren, würde man das Wesen des Mediendiskurses über die Arbeitslosigkeit verfehlen. Drittens ist die Fallzahl so gering, dass sich Testverfahren nicht sinnvoll anwenden lassen. Die Gründe sind, dass diese qualitativ-hermeneutischen Arbeiten so zeitaufwändig sind, dass nicht so viele Fälle analysiert werden können; sowie dass häufig historisch nur wenige Prozesse eines bestimmten Typs existieren.

## 6 Schlussfolgerungen für die Forschungspraxis

Ich habe vier Problembereiche quantitativer Längsschnittsanalysen herausgearbeitet, die Selbstbegrenzung der Forschergemeinschaft, Grenzen der bisherigen Auswertungsverfahren, ein Mangel an Daten sowie theoretische Grenzen. Was folgt nun aus diesen Problemen?

Eine typische Reaktion ist die theoretische Selbstbeschränkung: Der Analysezeitraum wird nicht auf das theoretisch sinnvolle sondern das Machbare beschränkt. Fragestellungen, die mit leicht zugänglichen Daten nicht beantwortet werden, werden ausgeschlossen. Diesen Weg halte ich persönlich für inakzeptabel. Vielmehr möchte ich im folgenden eine Reihe von Vorschlägen machen, wie man die momentanen Grenzen quantitativer Längsschnittsanalyse hinausschieben könnte. Beschreitet man diese Forschungspfade, stößt man möglicherweise am Ende tatsächlich an absolute Grenzen. Selbst wenn dem so wäre, liegt zwischen dieser Grenze und der momentanen Praxis ein riesiges Feld derzeit brachliegender Erkenntnismöglichkeiten.

Was also ist zu tun? Erstens lässt sich die quantitative Längsschnittsforschung selbst optimieren. Zweitens können die Schwächen quantitativer Sozialforschung zum Teil durch qualitative Sozialforschung ausgeglichen werden. In der Einleitung habe ich bereits erwähnt, dass die Unterschiede zwischen den beiden Verfahrenstypen fließend sind. Statt die Unterschiede zwischen ihnen zu betonen, sollten wir viel stärker nach Gemeinsamkeiten und Möglichkeiten suchen, qualitative und quantitative Forschung zu verbinden (*Peters (1998): IX; 8-9*). Qualitative und quantitative Forschung können phasenweise aufeinander folgen oder komplementär zu einander sein (*Erzberger / Kelle (1999)*). Beide Verfahrenstypen haben ihre Daseinsberechtigung. Abhängig vom Erkenntnisinteresse und der Datengrundlage eignet sich bisweilen das eine, bisweilen das andere Verfahren besser. Ich werde zeigen, wie diese beiden Modelle für Längsschnittsanalysen erweitert werden müssen und für welche zeitbezogenen Erkenntnisinteressen sie jeweils geeignet sind.

### 6.1 Optimierung des Prozesses quantitativer Sozialforschung

#### a) *Erweiterung des gängigen Modells des Forschungsprozesses um zusätzliche Phasen*

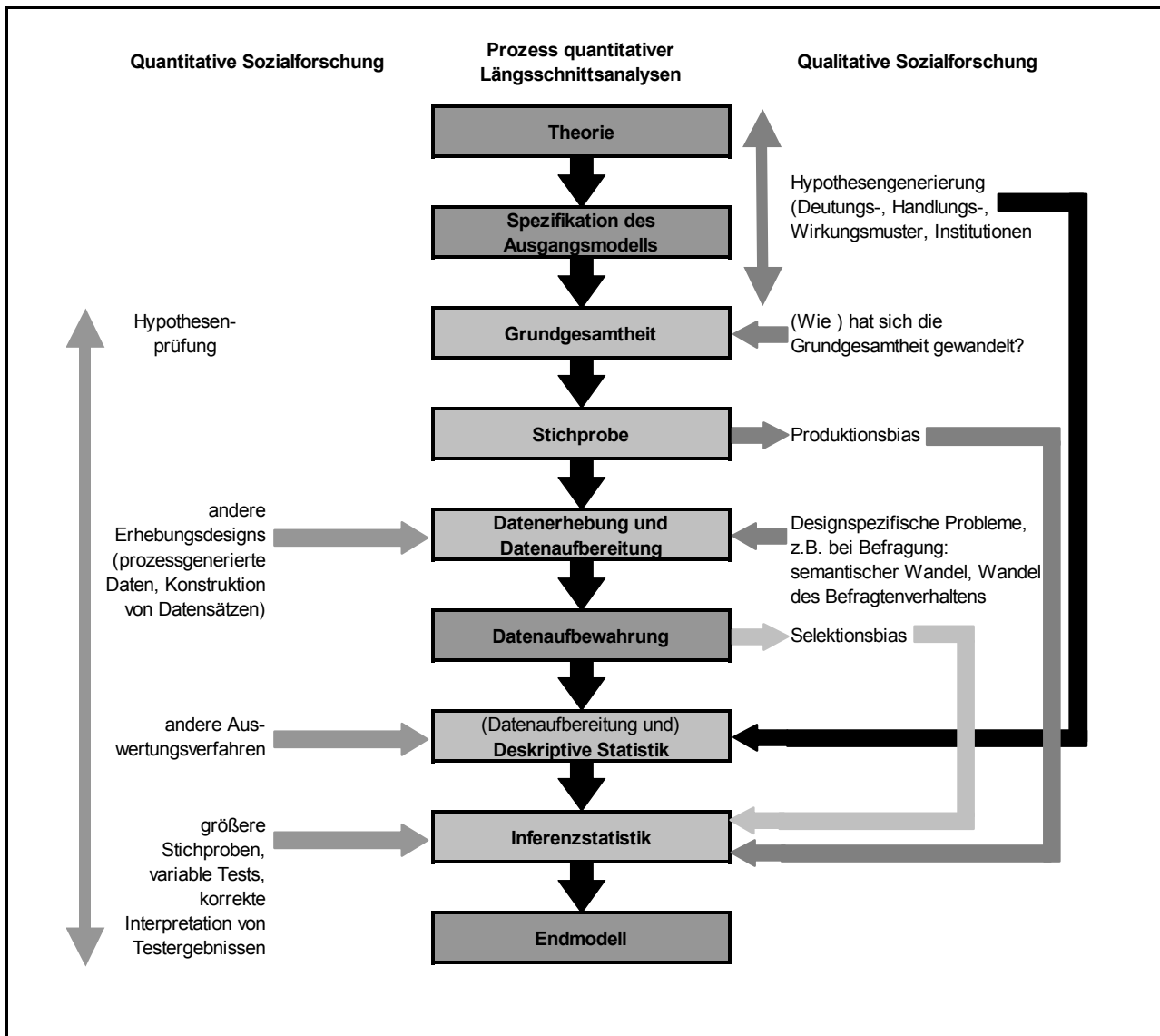
*Grafik 6* zeigt, wie der Prozess quantitativer Sozialforschung selbst optimiert werden kann. Ich schlage zunächst vor, das gängige Modell des Forschungsprozesses zu erweitern, und zwar in zwei Richtungen: Die erste Phase ist die Bildung einer Ausgangstheorie. Diese folgt die Spezifikation des Ausgangsmodells. Erst ausgehend von diesem kann die Grundgesamtheit festgelegt werden, aus der die Stichprobe gezogen wird. Am Ende des Forschungsprozesses steht ein präzisiertes, spezifiziertes oder modifiziertes Modells. Zwischen die Phasen der Datenerhebung und -auswertung sind zwei weitere Phasen einzuschieben: die Datenaufbereitung und die Datenaufbewahrung. In der Regel werden die Daten bei der Datenerhebung bereits zum ersten Mal aufbereitet, indem sie in den Datensatz eingelesen und Eingabefehler beseitigt werden. Häufig werden die Daten aber auch nochmals nach der Datenaufbewahrung, unmittelbar vor der Auswertung, aufbereitet, z. B. indem neue Variablen gebildet werden. Neben diesen zusätzlichen Phasen können einzelne Phasen des Forschungsprozesses optimiert werden.

#### b) *Datenerhebung und -aufbereitung*

Die meisten Forscher, die quantitative Längsschnittsanalysen betreiben, analysieren Daten, die im Trend- oder Paneldesign erhoben wurden. Es handelt sich hierbei meistens um Sekundäranalysen. Sekundärdaten sollten nicht benutzt werden, ohne vorher den Datenkonstruktionsprozess nachzuvollziehen oder zu überprüfen, ob die Daten zur Beantwortung der Forschungsfrage überhaupt geeignet sind. Dies gilt nicht nur für Individualsondern auch für Aggregatdaten, nicht nur für Rohdaten sondern auch für gebildete Skalen und Indizes.

Zwei Beispiele sollen verdeutlichen, was geschieht, wenn der Konstruktionsprozess an sich sauber konstruierter Indizes und Aggregatdaten nicht nachvollzogen wird. Das eine ist die Arbeitslosenquote. Diese wird in unterschiedlichen Ländern sehr unterschiedlich konstruiert.. (*Baur (1999, 2001)*). Immer wieder haben Regierungen die Arbeitslosenquote in ihrem Land „senken“ können, indem sie die Berechnungsgrundlage änderten. Das andere Beispiel ist der Preisindex: Jeder Verbraucher wusste nach der Euro-Umstellung, dass der Handel die Preise drastisch erhöht hatte. Der Preisindex verzeichnete jedoch keine außergewöhnliche Preissteigerung. Die Gründe hierfür bleiben verschleiert, wenn man den Konstruktionsprozess des Index sowie den Faktor „Zeit“ ausblendet. Zum einen spielte das Timing des Verhaltens der Hersteller eine Rolle: Zahlreiche Unternehmen hatten bereits im Vorfeld die Preise stark erhöht. Alternativ nahmen sie bestimmte – billigere – Produkte aus dem Handel und ersetzten sie durch neue Produkte, die entsprechend auch teurer waren. Letzteres zählt als Produktinnovation, nicht als Preissteigerung, auch wenn der Kunde lieber das alte Produkt gekauft hätte. Zum

**Grafik 6: Stufenmodell qualitativer und quantitativer Forschung**



anderen existiert ein Zielkonflikt zwischen gesellschaftlichem Wandel und Vergleichbarkeit statistischer Zahlen: Verschiedene Preisindizes sind nur vergleichbar, wenn der Warenkorb über die Zeit hinweg gleich bleibt. Was aber passiert, wenn die Menschen neue oder andere Produkte konsumieren? Passt man ihn nicht an, hat man einen Preisindex, der nichts mit den realen Konsummustern zu tun hat. Passt man ihn an – was das Statistische Bundesamt hin und wieder macht –, sind die Indizes zu verschiedenen Zeitpunkten nicht mehr vergleichbar. Anders ausgedrückt: Die Güter, die im Jahr 2000 im Warenkorb waren, entsprachen den realen Konsummustern vieler Verbraucher nicht und konnten die Preissteigerung so nicht erfassen.

Damit ein reflektierter Umgang mit Sekundärdaten möglich ist, müssen Primärforscher den gesamten Datenentstehungsprozess und die Mängel ihrer Datensätze besser dokumentieren, als dies derzeit üblich ist. Als Vorbild können dafür die in der qualitativen Sozialforschung entwickelten Standards dienen. Schnell (1997) fordert z. B. für standardisierte Befragungen ein Verlaufsprotokoll, das alle diese Entscheidungen (einschließlich Verlauf der Kontaktaufnahme mit dem Befragten) enthält.

Ein reflektierter Umgang mit häufig verwendeten Datensätzen genügt aber nicht. Ich habe oben ausgeführt, dass für eine bestimmte Forschungsfrage häufig keine Daten existieren, dass die existierenden Datensätze Mängel bei der Datenerhebung oder Stichprobenprobleme aufweisen. Will man verschiedene Subgruppen vergleichen, sind zudem häufig die Zeitfenster unterschiedlich. Eine bei weitem noch nicht ausgeschöpfte Möglichkeit, den Mangel an Daten zu bestimmten Themen zu umgehen, ist, auf andere Formen des Erhebungsdesigns zurückzugreifen. Insbesondere in drei Bereichen sehe ich noch ungenutzte Potentiale:

Erstens lässt sich Wandel der vergangenen fünfzig Jahre dadurch erfassen, dass man auf Replikationsstudien nach dem Muster von *Thomas Müller-Schneider* zurückgreift. Statt auf existierende Querschnittsdaten zurückzugreifen, können Forscher selbst Replikationsstudien durchführen.

Zweitens könnten quantitative prozessgenerierter Daten stärker genutzt werden. Dies gilt insbesondere für Firmendatenbanken und Datenbanken der öffentlichen Verwaltung. Hierbei stellt sich jedoch ein neues Problem: Häufig stehen diese Daten in völlig unterschiedlichen Datenbanken und Kodierungen zur Verfügung. Um auf sie zuzugreifen, muss zunächst ein Data Warehouse erstellt werden, so dass die Daten mit Data Mining-Techniken aufbereitet werden können. Dies würde aber bedeuten, dass Sozialwissenschaftler Data Mining-Kenntnisse erwerben müssten.<sup>20</sup>

Das größte Potential sehe ich allerdings – drittens – in der nachträglichen Konstruktion von Datensätzen. Dieses Verfahren erlaubt zumindest theoretisch die Konstruktion von Datensätzen für weit zurückliegende Zeiträume oder völlig neue Forschungsfragen. Zwei Beispiele habe ich bereits genannt: *Windolfs* Unternehmensdatenbank und unser Projekt zum Wandel des Diskurses über die Arbeitslosigkeit. Dieses Verfahren der Datenkonstruktion ist allerdings sehr aufwändig: Der Interpretationsaufwand ist sehr groß. Fundierte Kenntnisse im Bereich der qualitativen Sozialforschung und / oder der historischen Sozialforschung sind erforderlich. Ein häufiger Einwand gegen die nachträgliche Konstruktion von Datensätzen ist, dass sie weniger valide seien. Dies ist ein Irrtum: Zwar ist bei der Konstruktion solcher Datensätze äußerste Sorgfalt geboten – dieselbe Sorgfalt ist aber auch bei standardisierten Befragungen geboten. Dass der Datensatz von einem gut ausgebildeten Sozialforscher und nicht von einem „einfachen“ Befragten konstruiert wird, erhöht eher die Validität (*Baur / Behnke / Behnke* (2004); *Schulze* (2000)).

### c) Datenauswertung

Der Rückgriff auf andere Erhebungsdesigns beeinflusst die Auswertung: Spezifische Erhebungsdesigns implizieren *nicht* bestimmte Auswertungsverfahren und umgekehrt. Dennoch gehen bestimmte Auswertungsverfahren häufig mit bestimmten Erhebungsdesigns einher. Daraus folgt, dass – abhängig vom Typus verfügbarer Daten, variable Auswertungsstrategien gefahren werden müssen. Forscher müssen sich verschiedene Auswertungsverfahren beherrschen, um ihre Auswertungsstrategien der Forschungsfrage und den Daten anpassen zu können. Neben Menschen sollten vermehrt soziale Kollektive analysiert werden. Zudem sollten insbesondere in fünf Bereichen die existierenden Analyseverfahren weiterentwickelt werden: Benötigt werden Verfahren, die Stabilitäten identifizieren und die Verschränkung von Prozessen mit unterschiedlichem Wandlungstempo analysieren können. Mehrebenenanalysen und multivariate Verfahren müssen dynamisiert werden. Schließlich sollten die Sequenzanalyse und Optimal Matching-Techniken verstärkt eingesetzt werden, um Wandlungsmuster zu kennzeichnen.

Daneben muss der Umgang mit der Inferenzstatistik grundlegend überdacht werden: Standardmäßig sollten systematische Verzerrungen abgeschätzt werden. Testverfahren sollten nicht mechanisch angewandt sondern für das spezifische Erkenntnisinteresse entwickelt werden, d. h. Testverfahren müssten stärker variiert werden. Grundsätzlich sollte immer das *Gesamtmodell* getestet werden. In der Regel bedeutet dies, dass Stichproben größer sein müssen, als dies derzeit üblich ist. Schließlich ist darauf zu achten, dass Testergebnisse korrekt interpretiert werden. Insbesondere dürfen *Fisher-Tests* nicht mehr ständig als *Neyman-Pearson-Tests* fehl interpretiert werden.

Zum guten Teil sind diese Probleme bei der Datenauswertung und im Umgang mit Sekundärdaten längst bekannt. Dass sie dennoch übliche Praxis sind, ist zum Teil sicherlich Bequemlichkeit. Ein weiterer Grund liegt aber in der Methodenvermittlung: Auf viele dieser grundsätzlichen Fragen geht die Methodenausbildung nicht ein (*Ludwig-Mayerhofer* (2003)). Zudem existieren nach wie vor kaum gut verständliche Einführungen in fortgeschrittene statistische Verfahren. Dass wichtige Prinzipien der empirischen Sozialforschung systematisch und dauerhaft ignoriert werden, ist außerdem möglich, weil die Forschergemeinschaft dies durchgehen lässt. Immer wieder passieren Artikel den Peer-Review-Prozess, die Verfahren offensichtlich fehl interpretieren. Sollen diese Probleme behoben werden, muss sich diesbezüglich also auch die Gutachterpraxis ändern.

## 6.2 Validierung quantitativer durch qualitative Verfahren in bestimmten Forschungsphasen (Phasenmodell qualitativer und quantitativer Forschung)

Den Prozess quantitativer Längsschnittanalysen innerhalb der quantitativen Sozialforschung zu optimieren, genügt nicht, um alle Schwächen dieses Prozesses auszumerzen. Ich schlage deshalb vor, quantitative Verfahren in bestimmten Forschungsphasen durch qualitative Verfahren zu validieren. Dies entspricht dem Phasen-

<sup>20</sup> Zur Einführung siehe z. B. *Cabena / Hadjinian / Stadler / Verhees / Zanasi* (1997); *Knobloch / Weidner* (2000); *Knobloch* (2001); *Schur* (1994).



modell qualitativer und quantitativer Forschung nach *Erzberger* und *Kelle* (1999). Dieses ist in *Grafik 6* dargestellt.

#### **a) Hypothesengenerierung**

Bereits klassisch ist die Feststellung, dass qualitative Verfahren zur Hypothesengenerierung, quantitative zur Hypothesenprüfung dienen. Anders ausgedrückt: Mit Hilfe qualitativer Verfahren kann das Ausgangsmodell erstellt oder für den spezifischen Kontext modifiziert werden. Typische Deutungs-, Handlungs-, Interaktionsmuster können identifiziert werden und die Ausgestaltung und der Wandel des institutionellen Rahmens können nachgezeichnet werden. Dies erst ermöglicht die Spezifikation eines dem Kontext angemessenen Modells, das in der Auswertungsphase mit Hilfe quantitativer Verfahren überprüft werden kann.

#### **b) Wandel der Grundgesamtheit**

Doch auch in den Forschungsphasen, die typisch für die quantitative Sozialforschung sind, können punktuell angewandte qualitative Verfahren die Validität erhöhen. Dies beginnt bereits bei der Grundgesamtheit. Bevor diese festgelegt wird, sollte analysiert werden, ob sich diese gewandelt hat. Ist dies geschehen, ist zu analysieren, was dies für das Erkenntnisinteresse bedeutet und welche Stichprobe für die Grundgesamtheit repräsentativ ist. Wie ich oben dargestellt habe, müssen im Zweifelsfall verschiedene Analysen berechnet werden. Des Weiteren ist zu überlegen, ob die Stichprobe auch den Schluss auf die gewünschte Grundgesamtheit zulässt. Insbesondere die Gefahr ökologischer oder individualistischer Fehlschlüsse ist zu beachten. Werden verschiedene Gruppen (z. B. Länder) analysiert, ist zu überlegen, welche Zeitfenster miteinander verglichen werden sollten, um die Forschungsfrage zu beantworten. Ebenso ist zu prüfen, ob für diese gewünschten Analysezeiträume auch Daten existieren.

#### **c) Produktionsbias und Selektionsbias**

Bei der Datenproduktion und -aufbewahrung können systematische Fehler entstehen. Um abschätzen zu können, welcher Art diese sind und wie sie wirken, muss man den historischen und kulturellen Kontext genau kennen. Wenn über diesen nichts oder nur wenig bekannt ist, helfen nur qualitative Studien.

#### **d) Datenerhebung und -aufbereitung**

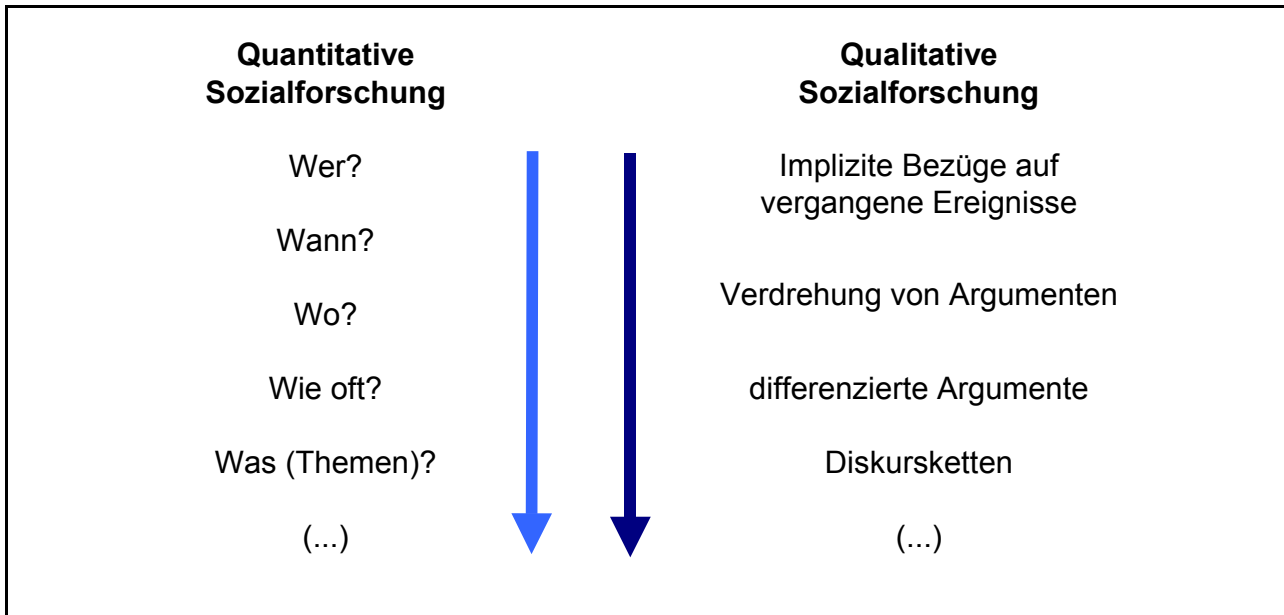
Wie qualitative Verfahren quantitative Verfahren in der Datenerhebungs- und -aufbereitungsphase ergänzen können, hängt vom spezifischen Designtyp ab. Bei standardisierten Befragungen ist z. B. zu fragen, ob sich das Befragtenverhalten gewandelt hat. Beispielsweise könnte ich mir vorstellen, dass das Thema Sexualität heute weniger oder zumindest auf andere Art und Weise tabuisiert ist als noch vor vierzig Jahren. Während man in den 1960ern über Geschlechtsverkehr grundsätzlich nicht redete, ist es heute vielleicht eher peinlich, keinen „guten Sex“ zu haben. Möglicherweise lügen Befragte heute auch mehr als früher usw. Diese Fragen beantworten kann man mit Hilfe qualitativer Verfahren. Semantische Analysen erlauben außerdem zu untersuchen, ob sich die semantische Bedeutung von Fragen oder bestimmten Worten gewandelt hat. Ganz anders sieht es Beispiel bei der nachträglichen Konstruktion von Datensätzen aus. Wie das Beispiel des Projekts „unempol“, das ich oben genannt habe, zeigt, ist der Datenerhebungsprozess an sich schon ein qualitatives Verfahren: Mit Hilfe inhaltsanalytischer Verfahren werden Aussagen klassifiziert und nachher kodiert.

### **6.3 Ergänzung quantitativer durch qualitative Forschungsergebnisse (Komplementaritätsmodell qualitativer und quantitativer Forschung)**

Mit Hilfe des Phasenmodells kann man also insbesondere Probleme im quantitativen Forschungsprozess und Probleme der Datengrundlage zumindest teilweise beheben. Ebenso kann man zwei der drei theoretischen Grenzen quantitativer Längsschnittsanalysen überwinden: Qualitative Vorstudien erlauben bei neuen Forschungsfragen und bei komplexen Kausalmodellen die Modellbildung und -spezifikation. Sie erlauben des Weiteren die Identifikation von Indikatoren, mit deren Hilfe diese Modelle quantitative überprüft werden können. Um die dritte Grenze – die Analyse von Deutungs- und Interaktionsmustern – zu überwinden, ist dagegen das Komplementaritätsmodell vonnöten.

Dies möchte ich am Beispiel des Projekts „unempol“ verdeutlichen (*Baglioni / della Porta / Graziano* (2004), *Baum / Lahusen* (2004), *Chabanet / Fay* (2004), *Linders* (2004), *Statham* (2004)): Wie bereits erwähnt, wurden in verschiedenen Ländern Zeitungsartikel zum Thema „Arbeitslosigkeit“ kodiert. Das Ergebnis ist ein quantitativer Datensatz, der Zeitvergleiche erlaubt. Wie *Grafik 7* zeigt, kann man mit Hilfe quantitativer Verfahren sehr

**Grafik 7: Komplementarität qualitativer und quantitativer Forschung am Beispiel des Diskurses über Arbeitslosigkeit**



gut nachzeichnen, welche Akteure sich wann und wo wie häufig geäußert haben und zu welchen Themen sie sich ungefähr geäußert haben. Man kann aufzeigen, ob sich diese Konstellationen im Lauf der Zeit gewandelt haben. Z. B. zeigt sich für Deutschland, dass im Jahr 1997 das Thema „Arbeitslosigkeit“ plötzlich wesentlich häufiger diskutiert wurde als zuvor und dass die Diskussion danach wieder abflaute. Vertreter der Opposition kommen wesentlich seltener zu Wort als Vertreter der jeweiligen Regierungspartei. Daneben diskutieren die Verbände stark mit, Arbeitslosenorganisationen und Privatpersonen kommen dagegen kaum zu Wort (*Baum / Lahusen* (2004)).

Eine Reihe von interessanten Fragen kann man mit Hilfe der quantitativen Daten aber *nicht* beantworten. Hierzu einige Beispiele: Der Diskurs bezieht sich häufig implizit auf vergangene Ereignisse, die nicht direkt zum Thema gehören, aber nicht genannt werden, weil sie allen Akteuren bekannt sind. Beispielsweise kam das Thema „Arbeitslosigkeit“ in Deutschland plötzlich in die Schlagzeilen, als die Zahl der Arbeitslosen erstmals über vier Millionen schnellte. Damit bei den Deutschen Erinnerungen an die Endzeit der Weimarer Republik wach, eine Tatsache, die sich verschließt, wenn man den Kontext nicht kennt. Ebenso wenig können die Daten differenzierte Argumente von Pauschalaussagen unterscheiden oder aufzeigen, inwiefern die Akteure aufeinander eingehen bzw. einander ignorieren. Mit Hilfe der quantitativen Daten kann man also keine Diskursketten nachzeichnen. Ebenso wenig legen sie offen, wie Argumente verstümmelt, verdreht und ins Gegenteil verkehrt werden, wie bestimmte Aussagen polemisiert werden usw. All dies muss qualitativ herausgearbeitet werden.<sup>21</sup> Bei solchen Forschungsfragen ergänzen sich also qualitative und quantitative Längsschnittanalysen gegenseitig.

## 7 Literatur

- Abbott, Andrew* (1995): **Sequence Analysis. New Methods and Old Ideas**. In: *Annual Review of Sociology*. Jg. 21. S. 93-113
- Abbott, Andrew / Tsay, Angela* (2000): **Sequence Analysis and Optimal Matching Methods in Sociology**. In: *Sociological Methods & Research*. Jg. 29. Heft 1. S. 3-33
- Abramson, Paul R. / Inglehart, Ronald* (1995): **Value Change in Global Perspective**. Ann Arbor: The University of Michigan Press
- Albert, Hans* (1999): **Die Soziologie und das Problem der Einheit der Wissenschaften**. In: *KZfSS*. Jg. 51. Heft 2. S. 215-231
- Albert, Hans* (2001): **Das Rahmenproblem und die disziplinäre Arbeitsteilung. Konsequenzen des methodologischen Revisionismus**. In: *KZfSS*. Jg. 53. Heft 4. S. 625-640
- Allmendinger, Jutta* (2002): **Eine drei-Welten-Lehre wissenschaftlicher Reputation und ihrer Messung**. In: *Soziologie*. Heft 3 / 2002. S. 56-58

<sup>21</sup> Man könnte nun argumentieren, dass man auch diese Aspekte verkodet und anschließend quantitativ analysieren könnte. Dies ist zwar richtig, aber die Hauptarbeit und auch das eigentlich Interessante ist dann, warum und wie man verkodet.

- Andreß, Hans-Jürgen (1985): **Multivariate Analyse von Verlaufsdaten. Statistische Grundlagen und Anwendungsbeispiele für die dynamische Analyse nicht-metrischer Merkmale.** Mannheim.
- Andreß, Hans-Jürgen / Heien, Thorsten / Hofäcker, Dirk (2001): **Wozu brauchen wir noch den Sozialstaat der deutsche Sozialstaat im Urteil seiner Bürger.** Wiesbaden: Westdeutscher Verlag
- Arnds, Pascal / Bonin, Holger (2002): **Frühverrentung in Deutschland: Ökonomische Anreize und institutionelle Strukturen.** Discussion Paper Nr. 666. Bonn: IZA
- Baglioni, Simone / della Porta, Donatella / Graziano, Paolo (2004): **Italy.** National Report on WP1 (first draft) on the project: unempol. The Contentious Politics of Unemployment in Europe. Political Claim-making, Policy Deliberation and Exclusion from the Labour Market
- Baum, Annerose / Lahusen, Christian (2004): **Germany.** National Report on Work-Package 1 on the project: unempol. The Contentious Politics of Unemployment in Europe. Political Claim-making, Policy Deliberation and Exclusion from the Labour Market
- Baur, Nina (1999): **Massenarbeitslosigkeit oder Arbeit für die Massen? Eine Analyse des Diskurses über die Zukunft der Erwerbsarbeit anhand ausgewählter soziologischer und ökonomischer Theorien.** Diplomarbeit im Studiengang Soziologie an der Otto-Friedrich-Universität Bamberg. <http://www.uni-bamberg.de/sowi/empirie/Baur/arbeit.pdf>
- Baur, Nina (2001): **Soziologische und ökonomische Theorien der Erwerbsarbeit. Eine Einführung.** Frankfurt a. M. / New York: Campus
- Baur, Nina (2003a): **Analyse zeitlicher Veränderungen mit Hilfe quantitativer Verfahren.** In: Baur, Nina (Hg.) *Quantitative Analyse zeitlicher Veränderungen.* Bd. 1: Überblick und theoretische Grundlagen. Reihe: Bamberger Beiträge zur empirischen Sozialforschung. S. 48-98
- Baur, Nina (2003b): **Bivariate Statistik, Drittvariablenkontrolle und das Ordinalskalensproblem. Eine Einführung in die Kausalanalyse und in den Umgang mit zweidimensionalen Häufigkeitsverteilungen mit SPSS für Windows.** Reihe: Bamberger Beiträge zur empirischen Sozialforschung. Bd. 9. Bamberg
- Baur, Nina (2003c): **Takeoff der Auswertung. Zur Vorbereitung statistischer Analysen.** Reihe: Bamberger Beiträge zur empirischen Sozialforschung. Bd. 5
- Beatty, John / Daston, Lorraine / Gigerenzer, Gerd / Krüger, Lorenz / Porter, Theodore / Swijtink, Zeno (1999): **Das Reich des Zufalls. Wissen zwischen Wahrscheinlichkeiten, Häufigkeiten und Unschärfen.** Heidelberg / Berlin: Spektrum Akademischer Verlag
- Beck, Ulrich (1986): **Risikogesellschaft. Auf dem Weg in eine andere Moderne.** Frankfurt am Main: Suhrkamp
- Beck, Ulrich (1994): **Jenseits von Stand und Klasse?** In: Beck, Ulrich / Beck-Gernsheim, Elisabeth (Hg.) (1994): *Risikante Freiheiten: Individualisierung in modernen Gesellschaften.* Frankfurt am Main: Suhrkamp. S. 43-60
- Beck-Bornholdt, Hans-Peter / Dubben, Hans-Hermann (1993a): **Der Hund, der Eier legt. Erkennen von Fehlinformationen durch Querdenken.** Reinbek: Rowohlt
- Beck-Bornholdt, Hans-Peter / Dubben, Hans-Hermann (2003b): **Der Schein der Weisen. Irrtümer und Fehlurteile im täglichen Denken.** Reinbek: Rowohlt
- Behnke, Joachim / Behnke, Nathalie / Baur, Nina (2004): **Empirische Methoden der Politikwissenschaft.** Paderborn: Ferdinand Schöningh. Im Erscheinen
- Beyer, Jürgen / Windolf, Paul (1995): **Kooperativer Kapitalismus – Unternehmensverflechtungen im internationalen Vergleich.** In: Nedelmann, Birgitta / Koepf, Thomas (Hg.) (1995): *Politische Institutionen im Wandel.* Opladen / Wiesbaden: Westdeutscher Verlag. S. 3-36
- Bloch, Marc (2002): **Apologie der Geschichtswissenschaft. Oder: Der Beruf des Historikers.** Stuttgart: Klett-Cotta
- Blossfeld, Hans-Peter / Hamerle, Alfred / Mayer, Karl Ulrich (1986): **Ereignisanalyse.** Frankfurt a. M.
- Blossfeld, Hans-Peter / Klijzing, Erik / Pohl, Katharina / Rohwer, Götz (1996): **Modellierung paralleler und interdependenter Prozesse in der Bevölkerungswissenschaft. Konzepte und Methoden am Beispiel der Heiratsneigung nichtehelicher Lebensgemeinschaften bei der Geburt des ersten Kindes.** In: *Zeitschrift für Bevölkerungswissenschaft.* Jg. 21. Heft 1. S. 29-56
- Blossfeld, Hans-Peter / Müller, Rolf (1996): **Sozialstrukturanalyse, Rational Choice Theorie und die Rolle der Zeit.** In: *Soziale Welt.* Bd. 47. S. 382-400
- Blossfeld, Hans-Peter / Prein, Gerald (1998): **The Relationship Between Rational Choice Theory and Large Scale Data Analysis. Past Developments and Future Perspectives.** In: Blossfeld, Hans-Peter / Prein, Gerald (Hg.) (1998): *Rational Choice Theory and Large-Scale Data Analysis.* Boulder: Westview. S. 3-27
- Blossfeld, Hans-Peter / Rohwer, Götz (1996): **Causal Inference, Time and Observation Plans in the Social Sciences.** Reihe: Sonderforschungsbereich 186 der Universität Bremen: Statuspassagen und Risikologen im Lebensverlauf. Arbeitspapier Nr. 36. Bremen
- Blossfeld, Hans-Peter / Rohwer, Götz (2002): **Techniques of Event History Modeling: New Approaches to Causal Analysis.** Mahwah (NJ): Erlbaum
- Blossfeld, Hans-Peter / Steinhage, Nikolei (1999): **Zur Problematik von Querschnittsdaten. Methodisch-statistische Beschränkungen bei der empirischen Überprüfung von Theorien.** Reihe: Globalife Working Paper Nr. 2 / 1999. Fakultät für Soziologie an der Universität Bielefeld. Bielefeld
- Bolte, Karl Martin (1961): **Einige Anmerkungen zur Problematik der Analyse von „Schichtungen“ in sozialen Systemen.** In: Glass, David W. / König, René (Hg.) (1968): *Soziale Schichtung und soziale Mobilität.* Köln / Opladen: Westdeutscher Verlag. S. 28-53
- Bolte, Karl Martin / Hradil, Stefan (1988): **Soziale Ungleichheit in der Bundesrepublik Deutschland.** Opladen: Leske + Budrich
- Braudel, Fernand (1950): **Standorte der Geschichte im Jahr 1950.** Braudel, Fernand (1992): *Schriften zur Geschichte 1.* Stuttgart: Klett-Cotta. S. 24-87
- Braudel, Fernand (1963): **Für eine serielle Geschichte: Sevilla und der Atlantik.** Braudel, Fernand (1992): *Schriften zur Geschichte 1.* Stuttgart: Klett-Cotta. S. 132-148
- Braudel, Fernand (1967): **Die Preise in Europa von 1450 bis 1750.** In: Braudel, Fernand (1993): *Schriften zur Geschichte 2.* Stuttgart: Klett-Cotta. S. 11-161
- Buchholz, Sandra / Hofäcker, Dirk (2003): **Late Careers and Career Exits in Times of Accelerating Social Change: Globalization and its Impact on Late-Midlife Careers. Conceptual Framework and German Case Study.** Präsentation im Rahmen des Doktorandenkolloquiums „Lebensverläufe im Globalisierungsprozess“. Otto-Friedrich-Universität Bamberg
- Cabena, Peter / Hadjinian, Peter / Stadler, Rolf / Verhees, Jaap / Zanasi, Alessandro (1997): **Discovering Data Mining. From Concept to Implementation.** Upper Saddle River (NJ): Prentice Hall

- Chabanet, Didier / Fay, Christophe (2004): **Preliminary Analysis – France (1995 – 2002)**. National Report (WP1) on the project: unempol. The Contentious Politics of Unemployment in Europe. Political Claim-making, Policy Deliberation and Exclusion from the Labour Market
- Daheim, Hansjürgen (1961): **Soziale Herkunft, Schule und Rekrutierung der Berufe**. In: Glass, David W. / König, René (Hg.) (1968): Soziale Schichtung und soziale Mobilität. 3. Auflage. Köln / Opladen: Westdeutscher Verlag. S. 200-217
- Daheim, Hansjürgen / Scheuch, Erwin K. (1961): **Sozialprestige und soziale Schichtung**. In: Glass, David W. / König, René (Hg.) (1968): Soziale Schichtung und soziale Mobilität. 3. Auflage. Köln / Opladen: Westdeutscher Verlag. S. 65-103
- Diekmann, Andreas (1988): **Ereignisanalyse – Beispiele, Probleme und Perspektiven**. In: ZUMA Nachrichten 23. S. 7-25
- Duby, Georges (1988b): **Der Sonntag von Bouvines. 27. Juli 1214**. Berlin: Wagenbach
- Endruweit, Günter (2002): **Wie misst man Reputation? Messtheoretische Überlegungen zur Jürgen Gerhards „Reputation in der deutschen Soziologie“**. In: Soziologie. Heft 4 / 2002. S. 33-41
- Endruweit, Günter (2003): **Definieren, messen und evaluieren. Antworten an Jürgen Gerhards**. In: Soziologie. Heft 1 / 2003. S. 110-111
- Engel, Uwe (1998): **Einführung in die Mehrebenenanalyse. Grundlagen, Auswertungsverfahren und praktische Beispiele**. Opladen: Westdeutscher Verlag
- Engel, Uwe (2002): **Methoden der empirischen Sozialforschung in Forschung und Lehre**. In: Soziologie. Forum der Deutschen Gesellschaft für Soziologie. Heft 2 / 2002. S. 78-89
- Erzberger, Christian (2001): **Sequenzmusteranalyse als fallorientierte Analysestrategie**. In: Sackmann, Reinhold / Wiggins, Matthias (Hg.) (2001): Strukturen des Lebenslaufs. Übergang, Sequenz, Verlauf. Weinheim: Juventa. S. 135-162
- Erzberger, Christian / Prein, Gerald (1997): **Optimal-Matching-Technik. Ein Analyseverfahren zur Vergleichbarkeit und Ordnung individuell differenter Lebensverläufe**. In: Zuma-Nachrichten. Jg. 21. Bd. 40. S. 52-80
- Erzberger, Christian / Kelle, Udo (1999): **Integration qualitativer und quantitativer Methoden. Methodologische Modelle und ihre Bedeutung für die Forschungspraxis**. In: KZfSS. Bd. 51. Heft 3. S. 509-531
- Esping-Andersen, Gøsta (1990): **The Three Worlds of Welfare Capitalism**. Cambridge (UK) / Oxford (UK): Polity / Blackwell
- Fromm, Sabine (2003): **Sequenzanalyse**. Working Paper. Otto-Friedrich-Universität Bamberg
- Gerhards, Jürgen (2002a): **Reputation in der deutschen Soziologie – zwei getrennte Welten**. In: Soziologie. Heft 2. Jg. 2002. S. 19-33
- Gerhards, Jürgen (2002b): **Zur Verbesserung der Selbstbeobachtung der Soziologie**. In: Soziologie. Heft 4. Jg. 2002. S. 56-65
- Gigerenzer, Gerd (1981): **Messung und Modellbildung in der Psychologie**. München / Basel: Ernst Reinhardt Verlag
- Gigerenzer, Gerd (1999): **Über den mechanischen Umgang mit statistischen Methoden**. In: Roth, Erwin / Holling, Heinz (Hg.) (1999): Sozialwissenschaftliche Methoden. Lehr- und Handbuch für Forschung und Praxis. 5. Auflage. München / Wien: R. Oldenbourg. S. 607-618
- Giugni, Marco / Statham, Paul (2002): **The Contentious Politics of Unemployment in Europe. Political Claim-Making, Policy Deliberation and Exclusion from the Labor Market. Codebook Workpackage 1: Political Claim-Making in the Public Domain**. Reihe: European Political Communication Working Paper Series. Bd. 2/02. Genua: University of Geneva. <http://www.leeds.ac.uk/ics/euro/unempol>
- Glenn, Norval D. (1977): **Cohort Analysis**. Newbury Park (CA) / London / New Delhi: Sage
- Grunow, Daniela (2004): **The Dynamics of Gender and Social Change: A Comparison of Male and Female Employment Careers in West Germany, East Germany and Denmark**. Präsentation im Rahmen des Doktorandenkolloquiums „Lebensverläufe im Globalisierungsprozess“. Otto-Friedrich-Universität Bamberg. 2.3.2004
- Hartung, Joachim / Elpelt, Bärbel / Klösener, Karl-Heinz (2002): **Statistik. Lehr- und Handbuch der angewandten Statistik**. 13., unwesentlich veränderte Auflage. München / Wien: R. Oldenbourg
- Hradil, Stefan (1999): **Soziale Ungleichheit in Deutschland**. 7. Auflage. Opladen: Leske + Budrich
- Huinink, Johannes (1988): **Methoden der explorativen Kohortenanalyse**. In: Zeitschrift für Bevölkerungswissenschaft. Jg. 14. Heft 1. S. 69-87
- Huinink, Johannes / Mayer, Karl Ulrich (1990): **Alters-, Perioden- und Kohorteneffekte in der Analyse von Lebensverläufen oder: Lexis ade?** In: Mayer, Karl Ulrich, (Hg.) (1990): Lebensverläufe und sozialer Wandel. S. 442-459
- Inglehart, Ronald (1989): **Kultureller Umbruch. Wertwandel in der westlichen Welt**. Frankfurt a. M. / New York: Campus
- Inglehart, Ronald (1997): **Modernization and Postmodernization. Cultural, Economic, and Political Change in 43 Societies**. Princeton (NJ): Princeton University Press
- Klüver, Jürgen (2003): **Gutachter hier, Gutachter da. Eine sehr persönliche Nachbemerkung zu einer metasozziologischen Diskussion**. In: Soziologie. Heft 1 / 2003. S. 112-117
- Knobloch, Bernd (2001): **Der Data-Mining-Ansatz zur Analyse betriebswirtschaftlicher Daten**. In: Informationssystemarchitekturen. Heft 8 (2001). S. 59-116
- Knobloch, Bernd / Weidner, Jens (2000): **Eine kritische Betrachtung von Data-Mining-Prozessen. Ablauf, Effizienz und Unterstützungspotentiale**. In: Jung, R. / Winter, R. (Hg.) (2000): Data Warehousing 2000. Methoden, Anwendungen, Strategien. Heidelberg: Physica. S. 345-365
- Koselleck, Reinhart (2000): **Zeitschichten. Studien zur Historik**. Frankfurt a. M.: Suhrkamp
- Lieberson, Stanley (1998): **Causal Analysis and Comparative Research. What Can we Learn from Studies Based on a Small Number of Cases?** In: Blossfeld, Hans-Peter / Prein, Gerald (Hg.) (1998): Rational Choice Theory and Large-Scale Data Analysis. Boulder: Westview. S. 129-145
- Linders, Annulla (2004): **Political Claim-making in the Public Domain (WP1). National Report Sweden**. Draft. Project: unempol. The Contentious Politics of Unemployment in Europe. Political Claim-making, Policy Deliberation and Exclusion from the Labour Market
- Lück, Detlev (2004): **Social Inequality and Social Justice: Do Modern Societies Accept More Differences?** Globalife Working Paper 44. Im Erscheinen
- Ludwig-Mayerhofer, Wolfgang (2003): **Zur Qualität der sozialwissenschaftlichen Methodenausbildung – am Beispiel statistischer Datenanalyse**. In: ZA-Informationen. Heft 53. S. 144-155

- Mayer, Karl Ulrich (1998): **Causality, Comparisons and Bad Practices in Empirical Social Research. A Comment on Stanley Lieberson's Chapter.** In: Blossfeld, Hans-Peter / Prein, Gerald (Hg.) (1998): *Rational Choice Theory and Large-Scale Data Analysis.* Boulder: Westview. S. 146-157
- Mayer, Karl Ulrich (1998): **Causality, Comparisons and Bad Practices in Empirical Social Research. A Comment on Stanley Lieberson's Chapter.** In: Blossfeld, Hans-Peter / Prein, Gerald (Hg.) (1998): *Rational Choice Theory and Large-Scale Data Analysis.* Boulder: Westview. S. 146-157
- Müller-Schneider, Thomas (1994): **Schichten und Erlebnismilieus. Der Wandel der Milieustruktur in der Bundesrepublik Deutschland.** Wiesbaden: Deutscher Universitätsverlag
- Müller-Schneider, Thomas (1996): **Wandel der Milieulandschaft in Deutschland. Von hierarchisierenden zu subjektorientierten Wahrnehmungsmustern.** In: *ZfS.* Jg. 25. Heft 3. S. 190-206
- Müller-Schneider, Thomas (2000): **Stabilität subjektzentrierter Strukturen. Das Lebensstilmodell von Schulze im Zeitvergleich.** In: *ZfS.* Jg. 29. Heft 5. S. 261-374
- Parsons, Talcott (1972): **Das System moderner Gesellschaften.** München: Juventa
- Parsons, Talcott (1975a): **Die Entstehung der Theorie des sozialen Systems: Ein Bericht zur Person.** In: Lazarsfeld, Paul F. / Parsons, Talcott / Shils, Edward (1975): *Soziologie – autobiographisch.* Stuttgart: Ferdinand Enke Verlag. S. 1-68
- Parsons, Talcott (1975b): **Gesellschaften. Evolutionäre und komparative Perspektiven.** Frankfurt am Main: Suhrkamp
- Peters, B. Guy (1998): **Comparative Politics. Theory and Methods.** Houndmills: Macmillan
- Rammert, Werner (2002): **Die halbierte Reputation – eine grob fahrlässige Rechnung!** In: *Soziologie.* Heft 3 / 2002. S. 53-55
- Roller, Edeltraud (2000): **Marktwirtschaftliche und wohlfahrtsstaatliche Gerechtigkeitsprinzipien in Deutschland und den USA.** In: Gerhards, Jürgen (Hg.) (2002): *Die Vermessung kultureller Unterschiede. USA und Deutschland im Vergleich.* Wiesbaden: Westdeutscher Verlag. S. 89-110
- Roller, Edeltraud (2001): **Wohlfahrtsstaat und Klassen Ende des sozialstaatlichen Konsenses.** In: *WZB-Mitteilungen.* Heft 93. S. 13-16
- Roller, Edeltraud (2002a): **Die Entwicklung der Akzeptanz des Sozialstaats und der Alterssicherung in Deutschland Mitte der siebziger Jahre.** In: *Deutsche Rentenversicherung.* Heft 9-10. Jg. 57. S. 510-522
- Roller, Edeltraud (2002b): **Erosion des sozialstaatlichen Konsenses und die Entstehung einer neuen Konfliktlinie in Deutschland.** In: *Aus Politik und Zeitgeschichte.* B 29/30. S. 13-19
- Rosenmayr, Leopold (1961): **Soziale Schichtung, Bildungsweg und Bildungsziel im Jugendalter.** In: Glass, David W. / König, René (Hg.) (1968): *Soziale Schichtung und soziale Mobilität.* Köln / Opladen: Westdeutscher Verlag. S. 268-283
- Schlittgen, Rainer / Streitberg, Bernd (2001): **Zeitreihenanalyse.** 9. Auflage. München / Wien
- Schnell, Rainer (1997): **Nonresponse in Bevölkerungsumfragen. Ausmaß, Entwicklung und Ursachen.** Opladen: Leske + Budrich
- Schulze, Gerhard (1996): **Die Erlebnisgesellschaft. Kultursociologie der Gegenwart.** 6. Auflage. Frankfurt a. M. / New York: Campus
- Schulze, Gerhard (2000): **Verstehen.** Paper 7 zum Hauptseminar „Wissenschaftstheorie für Sozialwissenschaftler“ an der Otto-Friedrich-Universität Bamberg im Wintersemester 2000 / 2001
- Schulze, Gerhard (2001a): **Welche Daten braucht die Soziologie?** Paper 1 zum Hauptseminar „Forschung und soziologische Theorie I“ an der Otto-Friedrich-Universität Bamberg im Wintersemester 2001 / 2002
- Schulze, Gerhard (2001b): **Theorie, Forschungshandeln, Methodologie. Über die Segmentierung von Argumentationsbereichen.** Paper 3 zum Hauptseminar „Soziologie der Forschung“ an der Otto-Friedrich-Universität Bamberg im Sommersemester 2001
- Schulze, Gerhard (2002a): **Einführung in die Methoden der empirischen Sozialforschung.** Reihe: Bamberger Beiträge zur empirischen Sozialforschung. Bd. 1. Bamberg: Otto-Friedrich-Universität Bamberg
- Schulze, Gerhard (2002b): **Standardisierung und Offenheit.** Paper 9 zum Hauptseminar „Forschung und soziologische Theorie II“ an der Otto-Friedrich-Universität Bamberg im Sommersemester 2002
- Schulze, Gerhard (2002c): **Offene Verfahren.** Paper 10 zum Hauptseminar „Forschung und soziologische Theorie II“ an der Otto-Friedrich-Universität Bamberg im Sommersemester 2002
- Schulze, Gerhard (2003a): **Die zeitliche Dimension sozialer Phänomene.** In: Baur, Nina (Hg.) (2003): *Quantitative Analyse zeitlicher Veränderungen. Eine anwendungsorientierte Einführung für Sozialwissenschaftler mit SPSS für Windows.* Bd. 1: Überblick und theoretische Grundlagen. Reihe: Bamberger Beiträge zur empirischen Sozialforschung. Bd. 19-1. Bamberg 2003. S. 8-27
- Schulze, Gerhard (2003b): **Erhebungsdesign und zeitliche Analyse.** In: Baur, Nina (Hg.) (2003): *Quantitative Analyse zeitlicher Veränderungen. Eine anwendungsorientierte Einführung für Sozialwissenschaftler mit SPSS für Windows.* Band 1. Reihe: Bamberger Beiträge zur empirischen Sozialforschung. Bd. 19-1. Bamberg 2003. S. 46-47
- Schulze, Gerhard (2004): **Individuelle und kollektive Merkmale.** Paper 11 zum Hauptseminar „Daten und Theorie I“ an der Otto-Friedrich-Universität Bamberg im Wintersemester 2002 / 2004
- Schur, Stephen G. (1994): **The Database Factory. Active Database for Enterprise Computing.** New York u. a.: John Wiley
- Srivastava, Jaideep (2004): **Mining Temporal Data.** <http://www.cs.umn.edu/research/websift/survey/sld001.htm> bis [sld037.htm](http://www.cs.umn.edu/research/websift/survey/sld037.htm) am 22.02.2004
- Statham, Paul (2004): **National Report on Political Claim-making Data-set for Britain.** First Draft. Project: unempol. The Contentious Politics of Unemployment in Europe. Political Claim-making, Policy Deliberation and Exclusion from the Labour Market
- Steinhage, Nikolei (2000): **Zeitaggregation und Ereignisdaten. Eine Simulation zu den Auswirkungen der Prozeßzeitkalierung.** Reihe: Globalife Working Paper Nr. 3 / 2000. Fakultät für Soziologie an der Universität Bielefeld. Bielefeld
- Windolf, Paul (2002): **Die Zukunft des Rheinischen Kapitalismus.** In: Allmendinger, Jutta / Hinz, Thomas (Hg.) (2002): *Organisationssoziologie. KZfSS. Sonderheft 42.* Opladen: Westdeutscher Verlag. S. 414-442
- Windolf, Paul (o. J.): **Forschungsprojekt Unternehmensverflechtung 1896 – 1938. Deutschland, USA und Frankreich im Vergleich.** Webadresse: [www.univier.de/uni/fb4/soziologie/apo/projektbeschreibung.htm](http://www.univier.de/uni/fb4/soziologie/apo/projektbeschreibung.htm) am 4.2.2004

