

Krütten, Rainer

**Book**

Synthetische Unternehmensstichproben: Ein alternativer Weg zur empirischen Analyse der Arbeitsnachfrage und deren Einfluß auf die Wertschöpfung im Unternehmen?

Organisationsökonomie humaner Dienstleistungen, No. 4

**Provided in Cooperation with:**

Rainer Hampp Verlag

*Suggested Citation:* Krütten, Rainer (1997) : Synthetische Unternehmensstichproben: Ein alternativer Weg zur empirischen Analyse der Arbeitsnachfrage und deren Einfluß auf die Wertschöpfung im Unternehmen?, Organisationsökonomie humaner Dienstleistungen, No. 4, ISBN 3-87988-255-X, Rainer Hampp Verlag, München und Mering

This Version is available at:

<https://hdl.handle.net/10419/116843>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

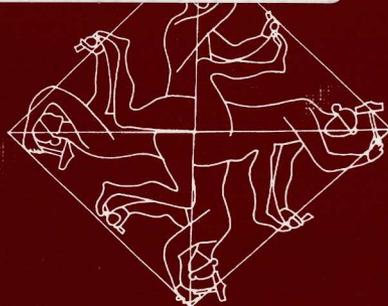
*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

**Organisationsökonomie  
humaner  
Dienstleistungen**

Rainer Krütten

**Synthetische Unternehmensstichproben**

A 216423



**Rainer Hampp Verlag**

EIGENTUM  
DES  
INSTITUTS  
FÜR  
WELTWIRTSCHAFT  
KIEL

---

BIBLIOTHEK

Synthetische Unternehmensstichproben.

# Organisationsökonomie humaner Dienstleistungen

Band 4

herausgegeben von Dieter Sadowski

Rainer Krütten

# Synthetische Unternehmensstichproben

Ein alternativer Weg  
zur empirischen Analyse der Arbeitsnachfrage und  
deren Einfluß auf die Wertschöpfung im Unternehmen?

776395

A 216423



Die Deutsche Bibliothek - CIP-Einheitsaufnahme

**Krütten, Rainer:**

Synthetische Unternehmensstichproben : ein alternativer Weg zur empirischen Analyse der Arbeitsnachfrage und deren Einfluß auf die Wertschöpfung im Unternehmen? / Rainer Krütten. - München ; Mering : Hampp, 1997

(Organisationsökonomie humaner Dienstleistungen ; 4)

Zugl.: Trier, Univ., Diss., 1996

ISBN 3-87988-255-X

Organisationsökonomie humaner Dienstleistungen: ISSN 0948-6518

Umschlagbild: Aus dem Skizzenbuch des Villard d'Honnecourt um 1235.  
Paris, Bibliothèque Nationale

*Liebe Leserinnen und Leser!*

*Wir wollen Ihnen ein gutes Buch liefern. Wenn Sie aus irgendwelchen Gründen nicht zufrieden sind, wenden Sie sich bitte an uns.*

© 1997 Rainer Hampp Verlag München und Mering  
Meringzeller Str. 16 D - 86415 Mering

Alle Rechte vorbehalten. Dieses Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne schriftliche Zustimmung des Verlags unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Mikroverfilmungen, Übersetzungen und die Einspeicherung in elektronische Systeme.

## GELEITWORT

Systematische empirische Arbeit in der Betriebswirtschaftslehre leidet sehr häufig unter dem Fehlen von betrieblichen Daten, wenn sie denn nicht mit finanzwirtschaftlichen Analysen von Jahresabschlüssen zu tun hat. Dies ist gerade für Analysen der betrieblichen Personalpolitik ein besonders großes Desiderat, die Empirielosigkeit der vorliegenden Lehrbücher ist dafür ein deutlicher Ausdruck.

Um den Mangel an Mikrodaten auszugleichen, gibt es heute zwei unterschiedliche Ansätze: Zum einen werden Unternehmenspanel, wie das NIFA-Panel deutscher Maschinenbauunternehmen, das Hannoveraner Panel oder das für Deutschland repräsentative IAB-Panel errichtet und mit bemerkenswertem Erfolg schon mikroökonomisch ausgewertet. Krütten gibt einen kurzen Überblick über die vorherrschenden Ansätze, konzentriert sich dann aber auf einen zweiten Weg, die Konstruktion von Unternehmensstichproben aus der Information, die aggregierte Randverteilungen enthalten. Es geht also gleichsam darum, aus statistischen Aggregaten auf die Populationsmitglieder zurückzuschließen. Es ist leicht vorstellbar, daß sehr unterschiedliche Populationen zu denselben Aggregationen führen - was man sich einfach daran veranschaulichen kann, daß ein und dieselbe durchschnittliche Unternehmensgröße für Stichproben sehr unterschiedlicher Unternehmen gelten kann: für sehr große und sehr kleine einerseits, gleich große andererseits etwa. Also ist das Auswahlkriterium für die Stichprobe, die man erreichen will, außerordentlich wichtig. Krütten übernimmt die in der einschlägigen Literatur begründete Vorstellung maximaler Entropie, die gleichbedeutend ist mit der Maßgabe, möglichst wenig zusätzliche Informationen in die Stichprobenkonstruktion zu stecken.

Daß die angewendeten Konstruktionsverfahren bekannte Stichproben recht gut in Mittelwert und Standardabweichung konstruieren, wie die Bonner Stichprobe deutscher publizitätspflichtiger Industrieunternehmen, ist ermutigend.

Krütten geht dann einen Schritt weiter und zeigt, daß mit seinen synthetischen Mikrodaten einige in der Literatur der Personalpolitik geäußerte Hypothesen, nämlich von B. Frick zur Beschäftigung Behinderter, anders zu beurteilen sind. Wenn Krütten auch nicht alle die Daten zur Verfügung standen, die für eine vollständige Kontrolle der Frickschen Hypothesen notwendig gewesen wären, so kann er doch im Prinzip einen Ausweg aus der Not, nämlich dem Mangel an Unternehmensdaten, veranschaulichen. Ob die Unentscheidbarkeit zwischen beiden Hypothesen dem Datenmangel geschuldet ist oder eine grundsätzliche Grenze synthetischer Unternehmensstichproben andeutet, bleibt unklar.

Das gleiche gilt für die spannende, von meinem Kollegen Wolfgang Sendler angeregte, weitreichende Frage, wie denn die Realität beschaffen sein muß, so daß die angewandten Konstruktionsverfahren oft zu so guten Approximationen der statistischen Kenngrößen führen. Es gibt einen Bedarf an weiterführender Forschung, der sich auf die beispielhaften Analysen Krüttens stützen kann.

Schloß Quint, Trier, im August 1997

Prof. Dr. Dieter Sadowski

## Vorwort

Entropiemaximierung ist das tragende Konstruktionsprinzip zur Erzeugung synthetischer Stichproben, die zur empirischen Überprüfung theoretischer Erkenntnisse auf dem Gebiet der ökonomischen Forschung herangezogen werden könnten. Dabei mißt die Entropie den Grad der Unbestimmtheit einer Verteilung, also auch den Grad der Ungewißheit für den Ausgang eines Versuchs. Um in der Terminologie der vorliegenden Arbeit zu bleiben, kann der Versuch dieser Promotion als lange mit hoher Entropie behaftet bezeichnet werden. Es waren sicherlich nicht nur die eher "allgemeinen" Risiken einer Promotion, die den Entstehungsprozeß dieser Arbeit in die Länge zogen, sondern auch die stark ausgeprägte "Ablenkungsfähigkeit" des Verfassers. So entwickelte sich sogar Rasenmähen zu einer der samstäglichen Lieblingsbeschäftigungen nur um den Leiden der Fertigstellung dieser Arbeit aus dem Wege zu gehen.

Wird bei der Konstruktion der synthetischen Stichprobe die Maximierung der Entropie angestrebt, so galt es, während des Entstehens dieser Arbeit die Entropie bezüglich des Ausgangs des Promotionsversuchs zu minimieren. Bei diesen Minimierungsbestrebungen unterstützte mich insbesondere meine Frau Brigitte, die mein fehlendes schlechtes Gewissen mehr als kompensierte, wenn ich mir nur allzu gerne eine Alternativbeschäftigung suchte. Für ihr wohl nicht immer leicht aufzubringendes Verständnis, den Verzicht auf viele "normale" Wochenenden und den trotzdem nie aufgegebenen Glauben an das Erreichen des Entropieminimums bezüglich der Fertigstellung der Arbeit danke ich ihr herzlich.

Wesentlich beteiligt am Prozeß der Entropie minimierung war auch mein Lehrer, Herr Prof. Dr. Dieter Sadowski, der mich während der Zeit als Student und Assistent an den spannenden Prozeß der wirtschaftswissenschaftlichen Forschung heranführte und so diese Arbeit erst ermöglichte. Seine Unterstützung und Förderung während dieser Zeit und auch darüberhinaus sowie seine große Geduld bis zur Fertigstellung der Arbeit waren sicherlich alles andere als selbstverständlich.

Meinen ehemaligen Kollegen Prof. Dr. Uschi Backes-Gellner und Prof. Dr. Bernd Frick als den wissenschaftlichen "Seniors" in einem Team junger Doktoranden gehört ebenfalls mein Dank, da sie mir stets als Gesprächspartner zur Verfügung standen und mich ein ums andere Mal nicht nur in die wissenschaftlich "richtige" Richtung lenkten.

Nicht vergessen möchte ich auch die anderen Mitglieder des Forschungsteams an der Universität und am Institut für Arbeitsrecht und Arbeitsbeziehungen in der EG (IAAEG), insbesondere Dr. Norbert Brühl und Dr. Kerstin Pull, die wesentlich zu einer dem Entstehungsprozeß dieser Arbeit förderlichen Kooperation beigetragen haben und schwierige Situationen zu meistern halfen.

Last but not least danke ich meinen Eltern, Elisabeth und Karl Krütten, ohne deren Vertrauen in den frühen Phasen meiner akademischen Ausbildung diese Arbeit wohl nicht entstanden wäre.

Trier, im Juli 1997

Rainer Krütten

# Inhaltsverzeichnis

Verzeichnis der Tabellen .....	13
Verzeichnis der Abbildungen .....	15
Verzeichnis der Übersichten .....	16
I. Einleitung: Problemstellung und Ziel der Arbeit .....	17
1. Datendefizit als eine mögliche Ursache für geringen Forschungsfortschritt .....	18
2. Die besondere Problematik von Unternehmensdaten .....	20
3. Konzeption und Vorgehensweise.....	21
II. Diskussion bisheriger Ansätze zur Gewinnung von Unternehmensdaten.....	25
1. Kriterien zur Bewertung unterschiedlicher Datensätze .....	25
2. Aggregat- vs. Mikrodaten .....	27
3. Diskussion und Beurteilung der einzelnen Datensätze .....	34
3.1 Aggregatdaten .....	34
3.1.1 Die amtliche Statistik .....	34
3.1.2 Daten der Berufsgenossenschaften.....	36
3.1.3 Die Betriebs- und Versichertendatei der Bundesanstalt für Arbeit .....	36
3.2 Unternehmensbefragungen und Panelerhebungen .....	37
3.3 Unternehmensdatenbanken .....	44
3.4 Fallstudien .....	46
III. Konstruktion synthetischer Daten mittels Simulated Annealing: Theoretische Analyse des Verfahrens und Experimente.....	47
1. Die Grundidee bei der Konstruktion synthetischer Mikrodaten mittels Simulated Annealing .....	47
2. Einordnung synthetischer Daten in den Beurteilungsrahmen.....	48
3. Theoretische Analyse der Konvergenzeigenschaften des Simulated Annealing .....	54

3.1	Definition des Optimierungsproblems .....	54
3.2	Das Prinzip der „Maximalen relativen Entropie“ .....	57
3.3	Simulated Annealing .....	61
3.3.1	Darstellung des Verfahrens .....	61
3.3.2	Konvergenz des Simulated Annealing zum globalen Kostenminimum .....	66
3.3.3	Kühlpläne .....	70
3.3.4	Kostenfunktion .....	75
3.3.5	Konvergenz gegen die „entropiemaximale Verteilung“ .....	77
3.3.6	Zusammenfassung der theoretischen Analyse des Simu- lated Annealing .....	79
4.	Ergebnisse kleinerer Experimente .....	81
4.1	Spezifikation der Parameter .....	82
4.2	Ergebnisse .....	83
5.	Rekonstruktion einer realen Stichprobe .....	99
6.	Zusammenfassung der Simulationsergebnisse .....	105
7.	Synthetische Daten und Gewichtung von Stichproben .....	106
IV.	Konstruktion einer synthetischen Unternehmensstichprobe zur Bestimmung der Determinanten der betrieblichen Schwerbehin- dertenbeschäftigung und deren Einfluß auf die Wertschöpfung im Unternehmen .....	109
1.	Vorbemerkung .....	109
2.	Theoretische Analyse der betrieblichen Nachfrage nach schwerbehinderten Arbeitnehmern .....	110
2.1	Betriebe als Akteure .....	112
2.2	Ziele der betrieblichen Personal- und Beschäftigungs- politik .....	112
2.3	Die Doppelfunktion interner Arbeitsmärkte .....	114
2.4	Zusammenfassung der Hypothesen .....	119
3.	Ausgewählte empirische Ergebnisse von Frick .....	121
3.1	Methodisches Vorgehen .....	121
3.2	Darstellung einzelner Ergebnisse .....	124

4.	Konstruktion der synthetischen Unternehmensstichprobe.....	127
4.1	Auswahl der Variablen und Aggregatinformationen .....	127
4.1.1	Die Auswahl der Aggregatinformationen .....	127
4.1.2	Die Auswahl der Variablen (Operationalisierung).....	129
4.2	Implementation des Simulated Annealing zur Generierung einer synthetischen Unternehmensstichprobe.....	137
4.2.1	Vorbemerkungen.....	137
4.2.2	Bestimmung der a priori-Verteilung und des Generierungsmechanismus .....	138
4.2.3	Spezifikation der übrigen Modellparameter.....	141
4.2.4	Darstellung des Programmablaufs.....	143
4.3	Ergebnisse der Simulation: Anpassung an die Randverteilungen .....	144
4.4	Empirische Ergebnisse zu den Bestimmungsgründen der betrieblichen Schwerbehindertenbeschäftigung und deren Einfluß auf die Wertschöpfung auf Basis der synthetischen Stichprobe .....	154
4.4.1	Vergleich der Segregationsindizes .....	154
4.4.2	Bestimmungsgründe der betrieblichen Arbeitsnachfrage: Produktions- und Faktornachfragefunktion .....	158
4.4.2.1	Theoretische Analyse: Produktionsfunktion und Faktornachfragefunktionen.....	159
4.4.2.2	Schätzung der Produktionsfunktion und Faktornachfragefunktionen .....	162
4.4.3	Bestimmungsgründe der betrieblichen Schwerbehindertenbeschäftigung.....	176
V.	Zusammenfassung und Ausblick .....	181
1.	Eignung des Simulated Annealing zur Gewinnung synthetischer Unternehmensdaten .....	181
2.	Wertschöpfung und Schwerbehindertenbeschäftigung: Zur empirischen Evidenz organisationsökonomischer Perspektiven auf Basis synthetischer Unternehmensdaten .....	183
3.	Weitere empirische Implikationen auf Basis synthetischer Unternehmensdaten.....	186

4. Ausblick: Zukünftige Forschung auf Basis synthetischer Unternehmensdaten.....	187
Anhang.....	191
Literaturverzeichnis .....	209

## Verzeichnis der Tabellen

Tabelle 1:	Standardabweichungen der jährlichen Veränderungs- raten von Neuzulassungen, Produktion und Beschäftig- tenzahl in der Automobilindustrie der Bundesrepublik Deutschland in verschiedenen Zeiträumen.....	30
Tabelle 2:	Schließungs- und Gründungs-raten von 1978-1985 und deren Einfluß auf den Beschäftigtenstand (alle Wirt- schaftszweige, ohne Landwirtschaft, Post, Bahn, Ge- bietskörperschaften und Sozialversicherung 1978- 1985).....	51
Tabelle 3:	Gemeinsame Verteilung zweier binärer Variablen bei vorgegebenen eindimensionalen Randverteilungen .....	78
Tabelle 4:	Vorgegebene paarweise Randverteilungen (Angaben in Prozent).....	81
Tabelle 5:	Entropiemaximale Verteilung zu den Randverteilungen in Tabelle 4.....	81
Tabelle 6:	Anfangsverteilung für einen Stichprobenumfang von 5.000 Fällen .....	82
Tabelle 7:	Zahl der kostenminimalen Zustände in Abhängigkeit vom Stichprobenumfang ( $L_k = n$ oder $5n$ ).....	85
Tabelle 8:	Anpassung an die entropiemaximale Verteilung für unterschiedliche Parameterkonstellationen .....	86
Tabelle 9:	Anpassung an die entropiemaximale Verteilung: Ag- gregation über alle Startwerte sowie über unterschied- liche Absenkungsgeschwindigkeiten.....	92
Tabelle 10:	Gefundene entropiemaximale Verteilungen für unter- schiedliche Parameterkonstellationen .....	84
Tabelle 11:	Vergleich der Korrelationskoeffizienten in der Bonner Stichprobe und der rekonstruierten Stichprobe .....	100
Tabelle 12:	Vergleich von Variablensumme, Mittelwert Standar- dabweichung, Minimum und Maximum in der jeweili- gen Größenklasse.....	101
Tabelle 13:	Anzahl der Betriebe nach Lohn- und Beschäftigtengrö- ßenklassen.....	104
Tabelle 14:	Anzahl der Betriebe nach Umsatz- und Beschäftigten- größenklassen .....	104

Tabelle 15:	Schwerbehindertenquote nach Wirtschaftsabteilungen und Betriebsgrößenklassen.....	111
Tabelle 16:	Vergleich des Anteils der weiblichen Beschäftigten auf Unternehmens- und Arbeitsstättenebene.....	133
Tabelle 17:	Vergleich der Zahl der Beschäftigten aus der Arbeitsstättenzählung und der DIW-Statistik auf der Ebene der Unterabteilungen für das Verarbeitende Gewerbe 1987 .....	136
Tabelle 18:	Relative Häufigkeiten und empirische Verteilungsfunktion der Unternehmen nach Betriebsgrößenklassen und Branchen.....	140
Tabelle 19:	Anpassung der Zahl der Unternehmen nach Größenklasse und Branche .....	147
Tabelle 20:	Anpassung der Zahl der Beschäftigten nach Größenklasse und Branche .....	148
Tabelle 21:	Anpassung für Durchschnittslohn nach Größenklasse und Branche.....	149
Tabelle 22:	Anpassung für Bruttoanlagevermögen pro Beschäftigtem nach Größenklasse und Branche .....	150
Tabelle 23:	Anpassung für Bruttowertschöpfung pro Beschäftigtem nach Größenklasse und Branche .....	151
Tabelle 24:	Anpassung Frauen, Facharbeiter, Azubis, Ausländer und Schwerbehinderte nach Branche .....	152
Tabelle 25:	Beschäftigungspflichtige Arbeitgeber ohne Schwerbehinderte.....	153
Tabelle 26:	Mittelwerte auf Unternehmens- und Branchenebene .....	154
Tabelle 27:	Segregationsindizes für Schwerbehinderte, Frauen und Ausländer.....	156
Tabelle 28:	Mittelwerte und Standardabweichungen der Variablen „Bruttowertschöpfung, Bruttoanlagevermögen und Beschäftigte“ sowie der logarithmierten Werte .....	164
Tabelle 29:	Cobb-Douglas-Produktionsfunktion für zwei Inputfaktoren.....	164
Tabelle 30:	Mittelwerte und Standardabweichung der Variablen „Frauenanteil, Facharbeiteranteil, Auszubildendenanteil, Schwerbehindertenanteil, Ausländeranteil und durchschnittliche Lohnhöhe pro Beschäftigtem.....	167
Tabelle 31:	Produktionsfunktion mit differenziertem Arbeitsinput .....	167

Tabelle 32:	Arbeitsnachfragefunktion für das Zwei-Faktoren-Modell.....	170
Tabelle 33:	Determinanten des Durchschnittslohns pro Beschäftigtem .....	174
Tabelle 34:	Arbeitsnachfragefunktion mit differenzierten erklärenden Variablen.....	175
Tabelle 35:	Nachfrage nach schwerbehinderten Arbeitnehmern .....	179
Tabelle 36:	Nachfrage nach schwerbehinderten Arbeitnehmern für Unternehmen mit beschäftigten Schwerbehinderten.....	180

## Verzeichnis der Abbildungen

Abbildung 1:	Produktion der sechs größten Unternehmen der Automobilindustrie der Bundesrepublik Deutschland 1970-1980 (1970 = 100) .....	32
Abbildung 2:	Anpassung an die entropiemaximale Verteilung in Abhängigkeit vom Stichprobenumfang für einen niedrigen Startwert des Kontrollparameters (0,1), hohe Absenkungsgeschwindigkeit (0,3) und Länge der Markovkette = n .....	93
Abbildung 3:	Mittelwert und Streuung der Kostenfunktionswerte für $n = 600$ ; $T_{\text{start}} = 10$ ; $a = 0,3$ und $L_k = n$ .....	94
Abbildung 4:	Durchschnittliche Akzeptanzwahrscheinlichkeit bei Verschlechterungen für $n = 600$ ; $T_{\text{start}} = 10$ ; $a = 0,3$ und $L_k = n$ .....	95
Abbildung 5:	Anpassung an die entropiemaximale Verteilung in Abhängigkeit vom Stichprobenumfang für einen niedrigen Startwert des Kontrollparameters (10), niedrige Absenkungsgeschwindigkeit (0,95) und Länge der Markovkette = $5n$ .....	96
Abbildung 6:	Mittelwerte und Streuung der Kostenfunktionswerte für $n = 600$ ; $T_{\text{start}} = 10$ ; $a = 0,95$ und $L_k = 5n$ .....	97
Abbildung 7:	Durchschnittliche Akzeptanzwahrscheinlichkeit bei Verschlechterungen für $n = 600$ ; $T_{\text{start}} = 10$ ; $a = 0,95$ und $L_k = 5n$ .....	98

Abbildung 8: Vergleich der Zahl der Beschäftigten aus der Arbeitsstättenzählung und der DIW-Statistik auf der Ebene der Unterabteilungen für das Verarbeitende Gewerbe 1987 .....	137
Abbildung 9: Absolute Differenz der Mittelwerte auf Unternehmens- und Branchenebene .....	154
Abbildung 10: Segregationsindizes für Schwerbehinderte, Frauen und Ausländer.....	157

## **Verzeichnis der Übersichten**

Übersicht 1: Kenndaten zu den Betriebspanel-Untersuchungen in den einzelnen Ländern.....	41
Übersicht 2: Vorgehensweise bei der empirischen Analyse von Frick.....	123

## **I. Einleitung: Problemstellung und Ziel der Arbeit**

Zur Analyse der Personal- und Beschäftigungspolitik von Unternehmen sind zum einen umfassende theoretische Analysen und zum anderen entsprechende empirische Überprüfungen der ausgearbeiteten theoretischen Konzepte notwendig. Die größten Defizite in diesem Analyseprozeß scheinen in der empirischen Arbeit zu liegen. Leontief (1982, zit. n. Stafford 1986, S. 387) formuliert in einem Brief an die Herausgeber der Zeitschrift "Science", daß die Mehrzahl der Artikel im American Economic Review bzw. der Ökonomie überhaupt empirisch falsch, trivial oder frei von Empirie sind, dagegen aber Theorie, Simulation und deplazierte Aggregation einen zentralen Platz einnehmen.

Für das Gebiet der Industrieökonomik zeichnet Scherer (1986, S. 3ff) ein ähnliches Bild: "Hunderte glänzender Theoretiker sitzen an ihren Schreibtischen und arbeiten an den neuesten Rätseln der Industrieökonomik, indem sie Kreisbahnen innerhalb weiterer Kreisbahnen aufzeichnen, um neue Modelle der Wirtschaft zu konstruieren." Ein echter Fortschritt wird aber kaum erzielt werden, solange es keine fruchtbare Zusammenarbeit zwischen denen gibt, "die das Zeug haben, immer genauere Beobachtungen der realen Welt durchzuführen und denen, die die empirischen Beweise in bessere theoretische Modelle von der Welt umsetzen können" (Scherer 1986, S. 4).

Die Ausführungen Scherers betonen eine starke Interdependenz zwischen theoretischem Erkenntnisfortschritt und entsprechenden Datengrundlagen. Den Argumenten Scherers folgend, ist eine Weiterentwicklung bewährten Wissens also nur dann zu erwarten, wenn aufgrund eingehender empirischer Tests nachgewiesen werden kann, daß Modelle die Realität gut abbilden oder bisherige theoretische Ansätze zu kurz greifen. Die vorliegende Arbeit hat es sich zum Ziel gesetzt, Defizite in dem für die Analyse der Personal- und Beschäftigungspolitik von Unternehmen vorhandenen Datenmaterial herauszuarbeiten und einen Weg aufzuzeigen, durch den für unterschiedliche Fragestellungen Daten bereitgestellt werden können.

Im Zusammenhang mit der Erklärung der Firmengrößen-Lohndifferentiale weisen Gerlach/Schmidt (1989, S. 370) darauf hin, daß ein erfolgversprechendes Forschungsprogramm erhebliche Anstrengungen und Neuorientierungen verlangt. Thaler (1989, S. 190) führt in einem Übersichtsartikel zur gleichen Problematik folgendes aus: "Economists would have to get their hands dirty collecting data on the actual operation of organizations. Unless the profession is willing to reward this type of time-consuming research activity, many important questions will remain unsolved."

Anhand umfangreicher empirischer Tests wird gezeigt, wie mit disaggregierten Daten Hypothesentests durchgeführt werden können und wie sich die Ergebnisse zu den Ergebnissen auf Basis von Aggregatdaten unterscheiden können. Die ver-

wendeten Daten wurden nicht originär erhoben, sondern mittels eines stochastischen Optimierungsverfahrens aus verschiedenen Aggregatdaten disaggregiert. Von Bedeutung ist also nicht nur die Verwendung von disaggregiertem Datenmaterial, sondern auch die Möglichkeit der Nutzung von Aggregatdaten aus unterschiedlichen Quellen.

### 1. Datendefizit als eine mögliche Ursache für geringen Forschungsfortschritt

Es besteht in der Literatur<sup>1</sup> ein weitgehender Konsens darüber, daß Datengewinnung auf einzelwirtschaftlicher Ebene eine lohnende Investition in die Theorie und Empirie der Wirtschaftswissenschaften ist<sup>2</sup>. Stafford (1986) und Hamermesh (1988) zeigen in ihren Studien, daß durch eine Bereitstellung zentraler Mikrodatensätze theoretisch fundierte empirische Forschungsarbeiten zu Themen, die sich mit den Datensätzen bearbeiten lassen, angeregt werden.

Stafford (1986, S. 386ff) untersucht 759 Arbeiten zur Arbeitsmarktforschung, die in sechs großen Zeitschriften in den USA zwischen 1965 und 1983 erschienen sind. Er zeigt, daß der Anteil der veröffentlichten Arbeiten zum Arbeitsangebot im Zeitraum von 1965-1979 stark zunahm, während der Anteil der Veröffentlichungen zur Arbeitsnachfrage im gleichen Zeitraum fast stagnierte. Für die Jahre 1980-1983 ist ein gegensätzlicher Trend zu beobachten. Die rückläufige Entwicklung bei den Arbeiten zum Arbeitsangebot setzt sich in den Jahren 1984-1987 fort (vgl. Hamermesh 1988, S. 9). Die Zahl der empirischen Arbeiten mit einer sinnvollen theoretischen Fundierung nimmt von 1965 bis 1983 stark zu. Die größte Zunahme ist bei den Arbeiten, die auf Mikrodatensätzen beruhen, zu verzeichnen, wohingegen die Zahl der Veröffentlichungen, die Zeitreihen nutzen, stark zurückgeht. Stafford führt den Anstieg der Arbeiten, die auf Mikrodaten beruhen, und der theoretisch fundierten empirischen Arbeiten auf die Bereitstellung entsprechender Mikrodatensätze zurück. Analyseeinheiten dieser Mikrodatensätze waren ausschließlich Individuen und Haushalte, so daß sich die starke Fixierung auf das Arbeitsangebot erklären läßt. Erklärungsbedürftig bleibt jedoch der Befund, daß trotz vorhandener Längsschnittdaten die Zahl der Arbeiten zum Arbeitsangebot von 1980-1987 zurückgeht. Stafford und Hamermesh formulieren übereinstimmend als Ursache für dieses Phänomen, daß die Erträge der wieder-

---

<sup>1</sup> Vgl. z. B.: Autoren-gemeinschaft 1990, S. 1ff; Fritsch 1990b, S. 411f.; Gerlach u. a. 1990, S. 1ff; Griliches 1986, S. 1468; Hamermesh 1988, S. 10ff; Hotchkiss 1990; Orcutt 1990, S. 30ff; Sadowski/Stengelhofen 1989, S. 105f.; Scherer 1986, S. 3ff; Stafford 1986, S. 387ff.

<sup>2</sup> Eine Ausnahme bilden hier die Vertreter eines instrumentalistischen Ansatzes, für die der empirische Wahrheitsgehalt von Hypothesen eine untergeordnete Rolle spielt, und die Prognosegüte eines Modells im Vordergrund steht (vgl. Fritsch 1990a, S. 24ff).

holten Nutzung einer Datenquelle abnehmend sind. "A reason for the decline in significant papers with an exclusively empirical orientation could be termed the data lode phenomenon. When a new type of data becomes available for research use as did micro cross-section and panel data in the 1970s, there appears to be a set of interesting results which will be based simply without any conceptual framework for interpretation. As the data lode is worked sufficiently, diminishing returns set in to purely data based findings and perhaps eventually to innovations in or reformulations of theory" (Stafford 1986, S. 394). Hamermesh führt aus: "I would argue, though, that at least in part it stems from the lack of new advances in the kinds (not the quantities) of data that are available for this purpose. The rich lode of the feedback relationship between data development and the expansion of knowledge about labor supply is now yielding decidedly diminishing returns" (Hamermesh 1988, S. 10).

Die Befunde der Autoren zeigen am Beispiel der Arbeiten zum Arbeitsangebot, wie durch ein Zusammenspiel von Theoriebildung, Methodenentwicklung und Datensammlung große Fortschritte im Bereich der Arbeitsmarktforschung erzielt werden können. "During the past twenty-five years research on the supply of labor has been the crown jewel of labor economics (perhaps even in all of applied economics). We have learned something about important phenomena; major threats of microeconomic theory that had not been used in empirical work have been explicitly employed in estimation; and these applications have generated important advances in econometric theory that have been used elsewhere in economics" (Hamermesh 1988, S. 6). Im Gegensatz dazu führen fehlende Datensätze auf Unternehmensebene neben kaum vorhandenen empirisch fundierten Arbeiten zur Arbeitsnachfrage zu Reformulierungen und Ausdehnungen in der Theoriebildung.

Für das Forschungsfeld der "Theorie der wirtschaftlichen Entwicklung" sieht Fritsch (1990a, S. 1f.) die Defizite vor allem darin begründet, daß Aggregatdaten auf Regional- oder Branchenebene nur beschränkte Aussagekraft besitzen. Ein wesentlicher Erkenntnisfortschritt sei nur dann zu erwarten, wenn die "Untersuchung eine Ebene "tiefer", nämlich bei den einzelnen Elementen dieser Aggregate, ansetzt" (Fritsch 1990a, S. 2), da Aggregate von Wirtschaftseinheiten wie Branchen, Größenklassen und Regionen wesentliche Entwicklungsdeterminanten nur unzureichend abbilden. Die Kenntnis der Determinanten einzelwirtschaftlichen Verhaltens ist nach Einschätzung des Autors deshalb von Bedeutung, weil einzelwirtschaftliche Einheiten die Adressaten wirtschaftspolitischer Maßnahmen darstellen.

Zu einer ähnlichen Aussage gelangen auch Hübler/Gerlach (1989, S. 172) im Zusammenhang mit der Überprüfung verschiedener Effizienzlohnmodelle auf Basis der ersten Welle des Sozio-ökonomischen Panels: "Bezüglich der Effizienzlohntheorie sei abschließend betont, daß die Domäne ihrer empirischen

Überprüfung die Ebene des Betriebes ist. Es ist deshalb erforderlich, die Effizienzlohnanalyse beruflicher Lohnstrukturen durch betriebliche Fallstudien und Paneldaten zu ergänzen."

Scherer (1986, S. 6ff) führt aus, daß in der industrieökonomischen Forschung ein lange vorherrschendes Paradigma durch die Analyse mittels Mikrodaten auf Unternehmensebene abgelöst wurde. Es konnte gezeigt werden, daß die Rentabilität bei größeren Marktanteilen kräftig steigt und im Gegensatz dazu die Zugehörigkeit zu einer hochkonzentrierten Industrie selbst im günstigsten Fall keinen Einfluß auf die Rentabilität hat. Auch in diesem Forschungsfeld, so Scherer, seien immer extensivere Theorien entwickelt worden; die Datensammlung zur empirischen Überprüfung der Theorien habe aber mit der Theorieentwicklung nicht Schritt gehalten.

## **2. Die besondere Problematik von Unternehmensdaten**

Die oben kurz wiedergegebene Diskussion zeigt eindrucksvoll den Zusammenhang zwischen Erkenntnisfortschritt und Datenverfügbarkeit. Für die Gewinnung geeigneter Daten auf Individualebene zur Beantwortung arbeitsangebotsseitiger Fragestellungen sind in den letzten Jahren erhebliche Mittel aufgewendet worden. Es ist jedoch ungewiß, ob die Bereitstellung entsprechender Mittel zur Gewinnung von Unternehmensdaten zum gleichen Erfolg führt. Die sehr kontroverse Diskussion über die Aufnahme der Jahresabschlüsse von kleinen und mittleren Kapitalgesellschaften, deren Firmendaten nur im Handelsregister hinterlegt werden, in Unternehmensdatenbanken zeigt, daß die Weitergabe von Daten durch die Unternehmen von diesen sehr skeptisch beurteilt wird (vgl. Börner 1988; Schmidt 1988).

Fritsch (1990a, S. 33f.) bedauert zwar, daß das reichhaltige Material der amtlichen Statistik unter anderem aufgrund restriktiv gehandhabter Datenschutzbestimmungen ein mit viel Aufwand erhobener "Daten-Friedhof" ist, dessen Potential für wissenschaftliche Untersuchungen kaum im Ansatz ausgeschöpft wird, verweist aber gleichzeitig auf die Vorteile von durch die Forscher selbst durchgeführten Primärerhebungen. Da die Weitergabe von Daten für Forschungszwecke durch die Unternehmen zum Teil sehr restriktiv gehandhabt wird, kommt es bei von den Forschern selbst durchgeführten Primärerhebungen zu einem unvollständigen Rücklauf, da im Gegensatz zur amtlichen Statistik keine Auskunftspflicht besteht. "Selbst bei einer relativ hohen Rücklaufquote entsteht so bei eigenen Primärerhebungen das Problem, daß man kaum vollständig sicher sein kann, inwiefern die zur Verfügung stehenden Daten nicht eine irgendwie verzerrte Stichprobe der Grundgesamtheit darstellen, und inwieweit die gewonnenen Ergebnisse letztendlich repräsentativ sind. Die in letzter Zeit zunehmende Zahl von derartigen Primärerhebungen führt außerdem dazu, daß immer mehr Betriebe

bzw. Unternehmen, welche grundsätzlich zu einer Auskunft bereit wären, sich als über Gebühr strapaziert empfinden und eine Teilnahme an derartigen Erhebungen neben der amtlichen Statistik ablehnen; sie erkennen letztlich, daß hier häufig im Prinzip vermeidbare Doppelarbeit geleistet wird" (Fritsch 1990a, S. 34). Auch bei Individualdaten stellt sich das Problem des unvollständigen Rücklaufs. Da aber die Struktur der Grundgesamtheit besser bekannt ist, kann eine Repräsentativität durch entsprechende Gewichtung erreicht werden. Zudem werden Personen nicht in dem Maße zu Befragungen durch die amtliche Statistik herangezogen wie Unternehmen.

Strebt man eine Wiederholungsbefragung (Betriebspanel) an, dann erhöht sich die Gefahr einer sinkenden Rücklaufquote neben einer in jeder Periode zu befürchtenden Antwortverweigerung noch durch die Schließung von Unternehmen bzw. deren Verschmelzung mit anderen Unternehmen. Unternehmensneugründungen können nur durch eine Ausdehnung der Stichprobe erfaßt werden<sup>3</sup>. Im Sozio-ökonomischen Panel werden Haushaltsaufspaltungen und -neugründungen durch die Weiterverfolgung bereits im Panel befragter Personen erfaßt.

### 3. Konzeption und Vorgehensweise

Nachdem in den vorherigen Abschnitten einige wesentliche Argumente für die Nutzung von Daten auf Mikroebene (Betriebe oder Unternehmen) vorgestellt worden sind, erscheint die Bereitstellung dieser Daten als gewichtiges Problem in der empirisch fundierten wissenschaftlichen Forschung. Im folgenden Kapitel II erfolgt daher eine exemplarische Bestandsaufnahme bisheriger Ansätze zur Beseitigung des Datenproblems. Dabei wird keine vollständige Auflistung aller verfügbaren Datenquellen angestrebt, sondern an Beispielen zu verschiedenen Typen von Datenquellen werden die Vor- und Nachteile bisheriger Ansätze diskutiert. Grundlage der Beurteilung bisheriger Ansätze bildet ein von Hamermesh (1988, S. 3ff) vorgeschlagener Anforderungskatalog für Datenquellen, die zur Analyse betrieblicher Personal- und Beschäftigungspolitik genutzt werden.

Aufbauend auf diesen Überlegungen wird in Kapitel III ein alternativer Weg der Datengewinnung auf einzelwirtschaftlicher Ebene vorgestellt. Vorhandene Aggregatdaten werden durch die Anwendung eines Algorithmus zur Lösung kombinatorischer Optimierungsprobleme, des "Simulated Annealing", disaggregiert. Dabei weicht die synthetisch erzeugte Verteilung nur soweit von einer Initialverteilung (meist Gleichverteilung) ab, wie die Randbedingungen dies erfordern.

Da diese Vorgehensweise sehr willkürlich erscheint, und die berechnete Frage gestellt werden kann, ob die synthetische Stichprobe ein Abbild der Grundge-

---

<sup>3</sup> Vgl. hierzu die Diskussion zum IAB-Betriebspanel in Abschnitt II.3.2.

samtheit ist, bedarf es einer ausführlichen Analyse des eingesetzten Verfahrens und der damit erzeugten Stichproben. Die Begründung für die Sinnhaftigkeit dieses Vorgehens stützt sich im wesentlichen auf drei Punkte:

### **1. Theoretische Analyse des eingesetzten Verfahrens:**

Da die "wahre" Verteilung der Variablen in der Grundgesamtheit unbekannt ist, muß eine Verteilung gefunden werden, die eine möglichst gute Schätzung der unbekanntenen Verteilung darstellt. Zunächst wird gezeigt, daß diejenige Verteilung, die die Entropie bezüglich einer beliebigen Ausgangsverteilung maximiert, die beste Schätzung für die unbekanntene Verteilung ist. Hierzu ist eine Diskussion des Konstruktionsprinzips "Maximierung der Entropie" erforderlich. In einem zweiten Schritt der theoretischen Analyse werden die Eigenschaften (Konvergenz in Abhängigkeit von Kostenfunktion und Kühlplan) des Simulated Annealing durch Anwendung der Theorie von Markovketten genauer untersucht.

### **2. Vergleich einer durch das Simulated Annealing erzeugten Verteilung mit bekannter entropiemaximaler Verteilung:**

Für eine kleine Zahl von Variablen und Ausprägungen kann die entropiemaximale Verteilung bezüglich vorgegebener Randbedingungen ermittelt werden. Ein Vergleich von synthetisch erzeugter und der entropiemaximalen Verteilung ermöglicht es zu überprüfen, ob die gewählte Implementation des Algorithmus Verteilungen mit den theoretisch postulierten Eigenschaften erzeugt.

### **3. Rekonstruktion einer realen Stichprobe:**

Aggregiert man die Daten einer realen Stichprobe und verwendet diese Daten als Grundlage für die Simulation, dann ist ein Vergleich von synthetischer und realer Stichprobe möglich. Es kann zudem eine Aussage über den Zusammenhang von verwendeten Randinformationen und der Qualität der Reproduktion gemacht werden.

Im Anschluß an diese Analysen des Simulated Annealing wird in Kapitel IV das Vorgehen bei der Konstruktion einer synthetischen Unternehmensstichprobe<sup>4</sup> detailliert dargestellt und die erzeugte Stichprobe genauer analysiert. In einem substanzwissenschaftlichen Teil wird die synthetische Stichprobe genutzt, um die von Frick (1992) aus organisationsökonomischen Theorien abgeleiteten Determinanten der betrieblichen Schwerbehindertenbeschäftigung mit Betriebsdaten empirisch zu testen. Angestrebt wird ein Vergleich der empirischen Ergebnisse, die auf Aggregatanalysen einerseits und auf Betriebsdaten andererseits beruhen. Da mehrere Datenquellen, die unterschiedliche Variablen enthalten, zur Erzeugung der synthetischen Stichprobe genutzt werden können, wird eine Analyse des

---

<sup>4</sup> Der Begriff "Stichprobe" wird hier nicht im Sinne der exakten mathematischen Definition verstanden.

Einflusses der betrieblichen Schwerbehindertenbeschäftigung auf die Arbeitsproduktivität möglich.

Im Schlußkapitel schließlich wird beurteilt, inwieweit sich synthetische Daten zur Produktion empirischen Wissens eignen, und ob die Verwendung von Unternehmensdaten anstelle von Aggregatdaten zu anderen Ergebnissen führt. Zudem soll ein Ausblick gegeben werden, wie durch Verwendung internationaler Statistiken synthetische Stichproben für andere Länder generiert werden können, um eine verbesserte Grundlage für international vergleichende empirische Forschung bereitzustellen.



## II. Diskussion bisheriger Ansätze zur Gewinnung von Unternehmensdaten

### 1. Kriterien zur Bewertung unterschiedlicher Datensätze<sup>5</sup>

Die Beurteilung, inwieweit sich bestimmte Datenquellen zum Testen theoretisch abgeleiteter Hypothesen eignen, soll vor dem Hintergrund eines von Hamermesh vorgeschlagenen Rahmens erfolgen. Ziel wirtschaftswissenschaftlicher Forschung ist häufig die Spezifikation und die explizite Bestimmung der Parameter eines allgemeinen linearen Modells der Form:

$$(1) \quad y_t = bx_t + e_t$$

mit:

y: abhängige Variable, x: Vektor der unabhängigen Variablen (theoretisch abgeleitet), b: Vektor von Parametern, die die Struktur des Kausalzusammenhangs kennzeichnen, t: Zeitpunkt, für den der Zusammenhang gelten soll und e: Störterm. Geschätzt wird jedoch nicht diese Beziehung sondern:

$$(2) \quad Y_{IT} = b_{IT}X_{IT} + e_{IT}$$

Der Index I steht für die Analyseeinheit, die den Kausalzusammenhang repräsentieren soll, und T kennzeichnet den Zeitpunkt der Beobachtung dieser Analyseeinheit.<sup>6</sup> Hamermesh wählt dieses einfache lineare Modell nur zu Veranschaulichungszwecken, weist aber gleichzeitig darauf hin, daß die Diskussion auch auf elaboriertere Modelle ausgedehnt werden kann. Es wird angenommen, daß zur Schätzung von (2) ein geeignetes Verfahren gewählt wird (beispielsweise ein Kleinst-Quadrate-Schätzer), und damit  $\hat{b}_{IT}$  ein unverzerrter Schätzer des Parameters  $b_{IT}$  für den spezifischen Datensatz  $\{Y_{IT}, X_{IT}\}$  ist, so daß gilt:

$$(3) \quad E(\hat{b}_{IT}|b_{IT}) = b_{IT} \quad ^7$$

Nimmt man an, daß  $\hat{b}_{IT}$  alle wünschenswerten statistischen Eigenschaften besitzt, dann kann der Wert dieses Parameters als beste Schätzung für eine Vielzahl möglicher Werte von  $b_{IT}$ , die auf möglichen Datensätzen  $\{Y_{IT}, X_{IT}\}$  beruhen,

---

<sup>5</sup> Vgl. zu diesem Abschnitt Hamermesh 1988, S. 3ff.

<sup>6</sup> t und T fallen dabei wohl in den wenigsten Fällen zusammen, da die Erhebung von Daten oft weit vor der Publikation von Forschungsergebnissen liegt. Eine Ausnahme bilden hier prozeßproduzierte Daten, die möglicherweise sofort nach deren Erfassung zur Verfügung stehen.

<sup>7</sup> E(x): Erwartungswert.

angesehen werden.<sup>8</sup> Zu der Verteilung in möglichen Datensätzen gehört demzufolge eine entsprechende Verteilung der  $\hat{b}_{IT}$ . Von besonderem Interesse ist jedoch die Frage, ob der Erwartungswert der  $\hat{b}_{IT}$ , gegeben den Datensatz  $\{Y_{IT}, X_{IT}\}$ , gleich dem "wahren" Wert  $b$ , der die Kausalbeziehung beschreibt, ist:

$$(4) \quad E(\hat{b}_{IT} | \{Y_{IT}, X_{IT}\}) = b.$$

Ob Gleichung (4) erfüllt ist, hängt dabei von vier wesentlichen Fragen ab:

**a) Meß- und Spezifikationsfehler:**

Bildet der erhobene Satz von Variablen  $\{Y, X\}$  die wahren Variablen  $\{y, x\}$  aus Gleichung (1) ab?<sup>9</sup>

**b) Repräsentativität:**

Ist die Stichprobe der gezogenen Analyseeinheiten (Individuen, Organisationen, Unternehmen etc.), die zur Schätzung von Gleichung (2) genutzt wird, repräsentativ bezüglich der Grundgesamtheit, für die der theoretische Wirkungszusammenhang postuliert wird? Wenn die Analyse von Gleichung (2) mehr als eine ökonomische Fallstudie darstellen soll, dann sollte das Sample von Analyseeinheiten sehr groß sein oder einige typische Analyseeinheiten beinhalten.

**c) Strukturelle Veränderungen:**

Ermöglichen die Zeitpunkte  $T$ , an denen die Daten erhoben wurden, eine Aussage über den Zusammenhang von  $y$  und  $x$  zu einem beliebigen Zeitpunkt  $t$ ?

Die Beantwortung dieser Frage hat zwei Dimensionen: Im einfacheren Fall muß geprüft werden, ob die geschätzten  $\hat{b}_{IT}$  die Struktur des Kausalzusammenhangs noch für den gegenwärtigen Zeitpunkt wiedergeben oder der strukturelle Wandel so groß war, daß die Struktur heute eine ganz andere ist als zum Erhebungszeitpunkt der Daten. Das Modell kann in diesem Falle noch richtig spezifiziert sein; die quantitativen Ausprägungen der  $\hat{b}_{IT}$  sind jedoch von denen zum Erhebungszeitpunkt verschieden.

<sup>8</sup> Mehrmaliges Ziehen einer Stichprobe mit jeweils gleicher Auswahlmethode (Zufallsauswahl) führt zu unterschiedlich aussehenden Stichproben, so daß die  $b_{IT}$  mehr oder weniger stark streuen.

<sup>9</sup> Wird beispielsweise in Gleichung (1) ein Zusammenhang zwischen Höhe des Einkommens und betriebsspezifischem Humankapital postuliert, und betriebsspezifisches Humankapital durch die Betriebszugehörigkeitsdauer gemessen, dann muß geprüft werden, ob die gemessene Größe die theoretische Größe adäquat abbildet. Die Frage von Meß- und Spezifikationsfehlern ist im Einzelfall zu überprüfen, und einer Beantwortung wird daher im folgenden nicht weiter nachgegangen.

Komplizierter erweist sich die Situation, wenn zusätzliche Einflußfaktoren zum Zeitpunkt  $t$  für den Zusammenhang von  $y$  und  $x$  relevant werden. Wenn die notwendigen Daten der zusätzlichen Einflußfaktoren im Datensatz enthalten sind, und der Datensatz für den Zeitpunkt  $t$  als noch repräsentativ betrachtet werden kann, dann läßt sich durch eine neue Schätzung die Struktur des Kausalzusammenhangs in  $t$  ermitteln. Sind die Daten im ursprünglichen Erhebungszeitpunkt jedoch nicht erhoben worden bzw. grundsätzlich nicht erhebbar, dann ist eine Aussage zur Struktur des Zusammenhangs in  $t$  nicht möglich.

**d) Aggregationsniveau:**

Sind die Analyseeinheiten (Individuen, Unternehmen, Wirtschaftszweige) bzw. die Zeitpunkte der Erhebung geeignet, den Zusammenhang von  $y$  und  $x$  abzubilden?

Zum einen geht es hier um eine adäquate zeitliche Aggregation und zum anderen um eine angemessene räumliche Aggregation. Werden zu große Analyseeinheiten gewählt, dann ist die Annahme eines linearen Zusammenhangs nicht gerechtfertigt, und die darauf aufbauenden Schätzwerte sind verzerrt. Bei hinreichend kleinen Analyseeinheiten ist eine Identifikation von Nichtlinearitäten möglich, wohingegen aber auch ein Aggregationsniveau denkbar ist, bei dem alle Nichtlinearitäten und Diskontinuitäten nivelliert werden.<sup>10</sup>

In vielen Fällen liegt aber das Interesse nicht in der Schätzung eines Strukturmodells, sondern im Testen von Hypothesen. Zu untersuchen ist dann nicht die Abbildungsgenauigkeit der wahren  $b$  durch  $\hat{b}_{IT}$ , sondern die Verteilung der  $\hat{b}_{IT}$  und der potentiell möglichen  $b_{IT}$ , deren Gestalt von der realisierten Stichprobe abhängt.<sup>11</sup>

## 2. Aggregat- vs. Mikrodaten

In der Theoriebildung der Wirtschaftswissenschaften wird zumeist der methodische Individualismus in den Vordergrund gestellt, d.h. es wird von den Eigenschaften und dem Verhalten einzelwirtschaftlicher Einheiten ausgegangen. Wie

---

<sup>10</sup> Legt man bei der Schätzung von Einkommensfunktionen (z.B. nach Mincer 1974) Individualdaten zugrunde, dann zeigt sich ein abnehmender Grenzertrag zusätzlicher Bildungsjahre. Dieser Effekt würde möglicherweise bei einer Schätzung mit aggregierten Daten (z.B. Altersgruppen) nicht mehr zu identifizieren sein.

<sup>11</sup> Zur Untersuchung der Streuung der  $\hat{b}_{IT}$  werden von den gängigen Statistiksoftwarepaketen Konfidenzintervalle berechnet. Streuen die Werte um den Wert null, dann ist selbst über die Richtung des Wirkungszusammenhangs kaum eine Aussage möglich. Ursachen für um null streuende Koeffizienten können sehr kleine Stichprobenumfänge, falsche Analyseebenen oder auch falsch spezifizierte Modelle sein (vgl. z.B.: Schneeweis 1990, S. 66ff).

oben kurz skizziert, ist im Gegensatz dazu für das Forschungsgebiet der Arbeitsnachfrage eine empirische Überprüfung auf Basis von Mikrodaten auf Unternehmens- oder Betriebsebene und damit auf Basis der einzelwirtschaftlichen Einheiten bisher kaum erfolgt. Die überwiegende Verwendung von Aggregatdaten ist im wesentlichen darauf zurückzuführen, daß Daten dieser Art (Branchen-, Regional- und Betriebsgrößenaggregate) insbesondere von amtlichen Stellen (Statistisches Bundesamt, Bundesanstalt für Arbeit, Berufsgenossenschaften; vgl. für eine detaillierte Darstellung z. B. Albach/Dahremöller 1986, Dahremöller 1987, S. 5ff, Dahremöller 1988, S. 30ff, Schumacher/Kistner 1989) aufbereitet werden und allgemein zugänglich sind. In der Literatur werden als weitere Gründe für die Verwendung von Aggregatdaten deren größere Überschaubarkeit und die bessere Aufdeckbarkeit von Regelmäßigkeiten gesehen. Zudem können Zusammenhänge, die auf einzelwirtschaftlicher Ebene vorhanden sind, auf Aggregatenebene verschwinden (vgl. Samuelson/Nordhaus 1985, S. 10f).

Die Eignung von Aggregatdaten für die Analyse einzelwirtschaftlichen Verhaltens muß angesichts einer Vielzahl von Argumenten und/oder empirischen Belegen bezweifelt werden. So werden beispielsweise bei der Beurteilung der wirtschaftlichen Entwicklung (Schaffung neuer Arbeitsplätze) häufig Aggregatdaten herangezogen, obwohl sich in den seltensten Fällen eindeutige Wachstums- oder Schrumpfungsbranchen identifizieren lassen (vgl. Fritsch 1990a, S. 9, Cramer 1987, S. 27ff). Es kann gezeigt werden, daß selbst bei einer sehr tiefen Untergliederung der Branchen hohe Streuungen der analysierten Größen zu beobachten sind. Diese Heterogenität auf Branchen-, Regional- oder Betriebsgrößenklassenebene macht einen Schluß vom Aggregat auf die einzelwirtschaftliche Ebene sehr problematisch.<sup>12</sup> Aggregatdaten eignen sich zu Aussagen über einzelwirtschaftliche Einheiten dann, wenn die vorhandenen Daten die Eigenschaften der Mikroeinheiten annähernd exakt abbilden. Fritsch (1990a, S. 13f.) kommt zu dem Schluß, daß mit zunehmender Heterogenität innerhalb der einzelnen Aggregate die Verallgemeinerbarkeit von Aussagen auf einzelwirtschaftlicher Ebene abnimmt und mit zunehmendem Aggregationsgrad die Heterogenität steigt (vgl. auch Erhard 1983, S. 375).<sup>13</sup> Dieser Sachverhalt gewinnt verstärkt an Bedeutung,

---

12 Es ist leicht einsehbar, daß ein bestimmtes arithmetisches Mittel das Ergebnis unendlich vieler Kombinationen von Merkmalsausprägungen sein kann. Ein Problem, das die Heterogenität, wenn auch kaum noch quantifizierbar, innerhalb der Branchen verstärkt, ist in der Zuordnung von Unternehmen zu den einzelnen Branchen zu sehen. Erhard (1983, S. 370ff) zeigt, daß mehr als die Hälfte der fachlichen Unternehmensteile (ca. 36.000) im Verarbeitenden Gewerbe auf nur 13.000 Unternehmen verteilt sind, ein Unternehmen damit im Durchschnitt über 2,8 fachliche Unternehmensteile verfügt. Eine eindeutige Zuordnung zu einer Branche wird dadurch erheblich erschwert.

13 Die Zuordnung zu einer höher aggregierten Gruppe macht den in Fußnote 8 erwähnten Zuordnungsfehler möglicherweise kleiner, da die fachlichen Unternehmensteile bei einem höheren Aggregationsgrad "richtig" eingruppiert werden.

wenn innerhalb der einzelnen Aggregate gegenläufige Entwicklungen stattfinden, die sich durch eine Aggregation im schlimmsten Fall auf null saldieren.<sup>14</sup>

Die Aggregation von Daten hat, sofern Heterogenität vorhanden ist, eine Reduktion der Varianz zur Folge, so daß selbst schlecht spezifizierte ökonomische Modelle zu hohen Varianzaufklärungen führen können. Umgekehrt ist jedoch auch der Fall denkbar, daß die Schätzung von Modellen, deren theoretische Herleitung vom Verhalten der Mikroeinheiten ausgeht, zu insignifikanten Ergebnissen führt, da die auf der Mikroebene vorhandenen Zusammenhänge auf der Aggregatebene nicht mehr vorhanden sind. Ökonomische Modelle, die auf der Basis von Aggregatdaten erstellt werden, weisen meist eine größere Varianzaufklärung auf als Modelle, denen Mikrodaten zugrunde gelegt werden.

Ein anderes wichtiges Argument bezüglich der Anwendung statistischer Verfahren resultiert aus der Konstruktion statistischer Testverfahren zum Testen alternativer Hypothesen. Da die Irrtumswahrscheinlichkeit von Testverfahren mit wachsendem Stichprobenumfang abnimmt, kann - bei identischer Struktur zweier Stichproben<sup>15</sup> - eine Hypothese auf Basis der kleinen Stichprobe abgelehnt und auf Basis einer größeren Stichprobe nicht mehr abgelehnt werden. Häufig wird die Zahl der Fälle durch Aggregation sogar soweit reduziert, daß eine Anwendung statistischer Testverfahren fragwürdig erscheint (Fritsch 1990a, S. 38).

An der Studie von Köhler/Sengenberger (Konjunktur und Personalanpassung: Betriebliche Beschäftigungspolitik in der deutschen und amerikanischen Automobilindustrie, 1983) läßt sich der oben formulierte Beurteilungsrahmen gut illustrieren.<sup>16</sup> Ziel dieser Studie ist die Analyse von Personalanpassungen in Abhängigkeit von Nachfrageschwankungen (Produktions- und Absatzschwankungen). Die Veränderungsrate der Beschäftigtenzahl und bezahlten Wochenstunden zum Zeitpunkt  $t$  wird erklärt durch Veränderungsrate der Produktion und des Absatzes in der Periode  $t$  sowie in den Vorperioden. Dabei wählen die Autoren unterschiedliche zeitliche (Monats-, Quartals- und Jahresdaten) und horizontale Aggregationsniveaus der Variablen. Die abhängige Variable, Veränderungsrate der Beschäftigtenzahl, wird einerseits differenziert für Männer und Frauen, andererseits für verschiedene Leistungsgruppen ausgewertet. Zusätzlich wird die Branche Straßenfahrzeugbau noch in verschiedene Teilindustrien

---

<sup>14</sup> Beispiele für unterschiedliche Entwicklungen innerhalb von Aggregaten liefern z.B. Dostal 1989, S. 193f.; Matzner et. al. 1988, S. 79ff.

<sup>15</sup> Es werden Aggregatdaten als Stichprobe kleineren Umfangs (Aggregat als repräsentatives Unternehmen) und Mikrodaten auf Unternehmensebene als größere Stichprobe betrachtet.

<sup>16</sup> Für die vorgenommenen Schätzungen werden zwar keine Mikrodaten verwendet, aber durch die unterschiedlichen Aggregationsgrade läßt sich der Einfluß der Aggregationsebene auf die Schätzergebnisse gut veranschaulichen. Bei den folgenden Ausführungen stehen nicht die inhaltlichen Argumente im Mittelpunkt, sondern methodische Gesichtspunkte.

(Kraftwagenbau, Teileindustrie) aufgespalten. Um Strukturveränderungen aufzuzeigen bzw. Nichtlinearitäten im Trend umgehen zu können, haben die Autoren den gesamten Untersuchungszeitraum 1950-1980 in Teilzeiträume zerlegt, in denen die Trends annähernd linear verlaufen (vgl. Köhler/Sengenberger 1983, S. 199f).

Meß- und Spezifikationsfehler können für die in dieser Studie verwendeten Daten vernachlässigt werden, da es sich bei den Variablen um "harte" Daten handelt, die zum Teil sogar den Jahresabschlüssen zu entnehmen sind. Probleme der Repräsentativität treten nur dann auf, wenn für einzelne Unternehmen, andere Wirtschaftszweige oder gar für die gesamte Industrie Schlußfolgerungen gezogen werden sollen. Es ist aber nicht möglich, die Ergebnisse für die Automobilindustrie der Bundesrepublik auf die Automobilindustrie eines anderen Landes zu übertragen. Die Autoren kommen für die amerikanische Automobilindustrie zu ganz anderen Ergebnissen. Erste Hinweise auf strukturelle Veränderungen im Beobachtungszeitraum erhält man, wenn die Streuungen der einzelnen Variablen im gesamten Zeitraum den Werten für unterschiedliche Teilzeiträume gegenübergestellt werden (vgl. Tabelle 1).

Tabelle 1: Standardabweichungen der jährlichen Veränderungsrate von Neuzulassungen, Produktion und Beschäftigtenzahl in der Automobilindustrie der Bundesrepublik Deutschland in verschiedenen Zeiträumen

	Jährliche Veränderungsrate			
	1951-80	1951-60	1961-70	1971-80
Neuzulassungen Kraftwagen	11,40	7,06	11,16	11,20
Produktion Kraftwagen	13,18	8,22	10,79	11,24
Beschäftigtenzahl insgesamt	5,61	5,99	4,79	3,90
Beschäftigtenzahl Arbeiter	5,98	5,89	5,46	4,00

Quelle: Köhler/Sengenberger 1983, S. 204

Wie Tabelle 1 zeigt, ist die Streuung der einzelnen Variablen in den betrachteten Zeiträumen zum Teil sehr unterschiedlich. Besonders auffallend ist der Wert der Standardabweichung für die Produktion Kraftwagen, der für den Gesamtzeitraum wesentlich höher liegt als in den Teilzeiträumen. Die hier gewählte Einteilung des Untersuchungszeitraums ist recht willkürlich. Sie macht jedoch eine erste Einschätzung über vorliegende Strukturveränderungen möglich. Vergleicht man zusätzlich die geschätzten Produktionselastizitäten der jährlichen Veränderungsrate der Beschäftigung für den Zeitraum 1950-1980 mit denen von 1971-1980, dann steigt der Einfluß der Vorperiode auf die Beschäftigungsanpassung stark an. Offenbar hat sich die Anpassungsgeschwindigkeit in der Bundesrepublik

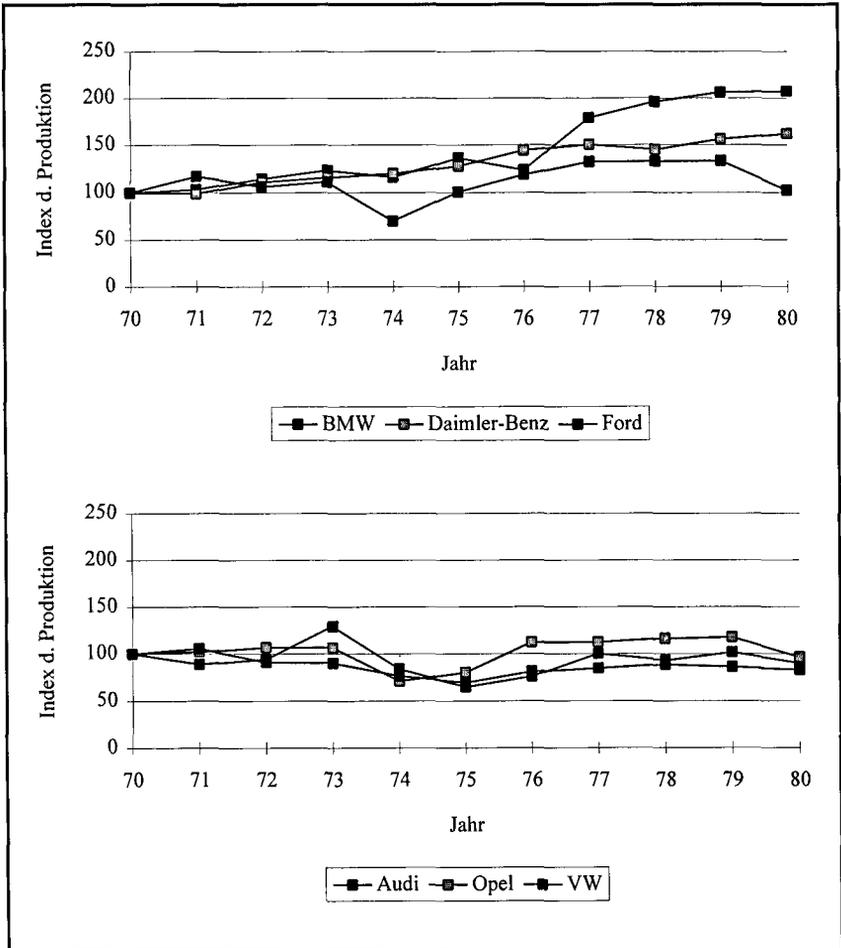
Deutschland im Untersuchungszeitraum verlangsamt. Im gesamten Untersuchungszeitraum hat die Änderungsrate der Produktion der aktuellen Periode und die der Vorperiode annähernd gleiche Bedeutung, während im Zeitraum von 1971-80 die relative Bedeutung der Produktionsveränderung in der Vorperiode zunimmt. Diese kurzen Überlegungen zeigen, daß die Länge der zugrundegelegten Zeitreihe die Ergebnisse ganz erheblich verändern kann. Es ist für den jeweiligen Untersuchungsgegenstand zu klären, welche Beobachtungszeiträume die heutige Struktur abbilden.

Die Wahl der angemessenen zeitlichen und räumlichen Aggregationsebene sowie deren Einfluß auf die Schätzergebnisse läßt sich an dieser Studie ebenfalls verdeutlichen. Bei der Schätzung von Elastizitäten der Beschäftigung verwenden die Autoren Aggregatdaten für die gesamte Automobilindustrie bzw. differenzieren für Kraftwagenbau und Teileindustrie. Die Entwicklung der Produktion in den sechs größten Unternehmen der Automobilindustrie zeigt für die Bundesrepublik von 1970-1980 ein sehr heterogenes Bild (vgl. Abbildung 1). Während einige Unternehmen (BMW, Daimler Benz), abgesehen von kleinen Produktionsrückgängen, ein stetiges Wachstum der Produktion erreichten, war die Entwicklung bei anderen sehr unterschiedlich. Zum Teil erreichte die Produktion 1980 den Stand von 1970. Bei den Beschäftigtenzahlen zeigt sich für den Untersuchungszeitraum ein ähnlich heterogenes Bild (vgl. Köhler/Sengenberger 1983, S. 212). Eine Interpretation der geschätzten Elastizitäten erklärt folglich nicht in jedem Falle die Entwicklung des einzelnen Unternehmens. Diese Hinweise auf erhebliche Heterogenität innerhalb der Aggregate werden bei genauerer Betrachtung der Schätzergebnisse verstärkt. Wenn die Schätzung der Produktionselastizität der Zahl der Beschäftigten (Arbeiter) für den Zeitraum 1970-1980 differenziert für Kraftwagenbau und Teileindustrie erfolgt, dann sinkt die Varianzaufklärung von 0,72 für die Gesamtindustrie auf 0,58 für die Teileindustrie (Jahresdaten). Werden zusätzlich noch mögliche strukturelle Veränderungen im Analysezeitraum ignoriert, und erfolgt eine Schätzung für den Zeitraum von 1956-1980, dann reduziert sich die Varianzaufklärung in der Teileindustrie für die Gruppe der Arbeiter sogar auf 0,26. Eine Aufspaltung der Beschäftigtenzahl nach Geschlecht und Lohngruppen führt generell zu einer geringeren Varianzaufklärung und zum Teil sogar zu insignifikanten Regressionskoeffizienten.

Die Verwendung verschiedener zeitlicher Aggregationsniveaus hat nicht unerheblichen Einfluß auf die Ergebnisse der Schätzungen. Für die Bundesrepublik ergeben sich bei Schätzungen auf Basis von Monats- oder Quartalsdaten sehr geringe Varianzaufklärungen und bei Monatsdaten zum Teil insignifikante Regressionskoeffizienten. Die besten Ergebnisse werden auf Basis der jährlichen Veränderungsdaten erzielt. Dieser Befund ist nicht notwendigerweise auf eine mit höherem Aggregationsniveau meist zunehmende Varianzaufklärung zurückzuführen, denn die hohen t-Werte der Regressionskoeffizienten deuten darauf hin, daß die jährlichen Veränderungsdaten den Zusammenhang am besten abbilden.

Für die amerikanische Automobilindustrie sind die Varianzaufklärungen bei Verwendung von Monats- und Quartalsdaten wesentlich höher als in der Bundesrepublik. Diese Ergebnisse können dahingehend interpretiert werden, daß die Wahl des räumlichen und/oder des zeitlichen Aggregationsniveaus großen Einfluß auf die Ergebnisse besitzt. Vernachlässigt man zusätzlich noch mögliche strukturelle Veränderungen im Zeitablauf, sind größere Schätzfehler zu erwarten.

Abbildung 1: Produktion der sechs größten Unternehmen der Automobilindustrie der Bundesrepublik Deutschland 1970-1980 (1970=100)



Quelle: eigene Erstellung, Daten aus Köhler/Sengenberger 1983, S. 212.

Die Autoren waren aufgrund des umfangreichen Datenmaterials in der Lage, viele Varianten des spezifizierten Modells zu schätzen und den Einfluß unterschiedlicher Daten auf die Schätzparameter zu testen. Es bleibt zu prüfen, ob bei einer Analyse auf Basis der originären Unternehmensdaten andere Ergebnisse erzielt werden. Da in dieser Studie sekundärstatistisches Material ausgewertet wurde, war eine Berücksichtigung unterschiedlicher Zeiträume und diverser zeitlicher und räumlicher Aggregationsniveaus möglich. Die Durchführung einer zur Beantwortung der Forschungsfrage geeigneten Unternehmensbefragung hätte alle oben skizzierten Probleme antizipieren müssen, da Nacherhebungen sehr teuer und zeitintensiv sind. Im Idealfall müßten die Daten in langen Zeitreihen und räumlich und zeitlich so disaggregiert wie möglich auf Unternehmensebene erhoben werden, so daß beliebige Aggregationen zu einem späteren Zeitpunkt vorgenommen werden können.

Bereits in einer sehr frühen Studie untersuchen Orcutt et al. (1968) den Zusammenhang zwischen Aggregationsniveau und Schätzergebnissen. Wegen fehlender Mikrodaten benutzen sie zur Schätzung eines makroökonomischen Modells einer Volkswirtschaft durch Simulation erzeugte Daten, die auf unterschiedlichen Niveaus aggregiert werden. Zudem werden unterschiedlich lange Zeitreihen erzeugt. Die Ergebnisse zeigen deutlich, daß die Benutzung gering aggregierter Daten zu weniger verzerrten Schätzern führt, die Präzision der Schätzergebnisse zunimmt, und Fehlspezifikationen in den Modellen identifiziert werden können (vgl. Orcutt et. al. 1968, S. 785).

Die wichtigsten Vorteile eines Ansatzes auf Mikroebene lassen sich wie folgt zusammenfassen (vgl. Fritsch 1990a, S. 40ff):

- tatsächlich vorhandene Heterogenität kann entdeckt und untersucht werden,
- eine Überprüfung, ob Aggregationsmerkmale aussagefähig sind, wird möglich,
- und Wirkungszusammenhänge können besser aufgedeckt werden, d.h. es können aus einzelwirtschaftlichen Untersuchungen Tests auf Fehlspezifikationen für Modelle, die auf Aggregatdaten aufbauen, hergeleitet werden.

Obwohl oben die Unterlegenheit eines aggregatanalytischen Ansatzes gegenüber einer einzelwirtschaftlichen Fundierung begründet wird, sprechen die leichte Zugänglichkeit und die erschöpfende Erhebung durch die entsprechenden Institutionen (zum Teil besteht Auskunftspflicht) für die Verwendung dieser Aggregatdaten. Diesem Datenmaterial liegen bei den Unternehmen originär erhobene Daten zugrunde, die aus datenschutzrechtlichen Gründen nicht weitergegeben werden dürfen. In diesem Projekt werden Aggregatdaten als Grundlage für die Konstruktion einer synthetischen Unternehmensstichprobe genutzt.

### **3. Diskussion und Beurteilung der einzelnen Datensätze<sup>17</sup>**

#### **3.1 Aggregatdaten**

Sekundärstatistisches Datenmaterial ist insbesondere vom Institut für Mittstandsforshung auf seine Eignung für den Aufbau einer Querschnitts-, Existenzgründungs- und Verlaufsstatistik für alle Wirtschaftsbereiche untersucht worden (vgl. Dahremöller 1988; Albach/Dahremöller 1986). Im Mittelpunkt der Analyse steht bei diesen Arbeiten eine möglichst genaue Erfassung des Gründungs- und Fluktuationsgeschehens von Unternehmen. Eine aktuelle Unternehmensbestandsstatistik würde eine Überprüfung der Repräsentativität anderer Unternehmensdaten ermöglichen.

##### **3.1.1 Die amtliche Statistik**

Unternehmens- und Betriebsdaten werden von der Amtlichen Statistik in der Arbeitsstättenzählung (zuletzt 1987, davor 1970), der Erhebung der industriellen Kleinbetriebe (industrielle Betriebe von Unternehmen mit unter 20 Beschäftigten ohne Handwerk), der monatlichen Erhebung im Handel und Gastgewerbe auf Stichprobenbasis, der Handels- und Gaststättenzählung (alle 5-10 Jahre), in der Monatsberichterstattung im Bergbau und Verarbeitenden Gewerbe (Betriebe mit 20 und mehr Beschäftigten) und der Umsatzsteuerstatistik (alle 2 Jahre) zur Verfügung gestellt.

Die umfassendste Erhebung ist die Volks-, Berufs- und Arbeitsstättenzählung, die in der Regel alle 10-15 Jahre durchgeführt wird. Wegen der großen Zeitabstände zwischen den einzelnen Erhebungen und der Zeitdauer von 2-3 Jahren bis zur Veröffentlichung der Daten eignet sich diese Datenquelle nur bedingt für die Ermittlung des aktuellen Unternehmensbestandes. Um diesen zu ermitteln, sind zeitliche Fortschreibungen notwendig, die jedoch eine Kenntnis des Gründungs- und Fluktuationsgeschehens voraussetzen. Ähnliches gilt für die Handels- und Gaststättenzählung, die als Totalerhebung alle 10 Jahre durchgeführt wird (vgl. Dahremöller 1988, S. 30).

Relativ gut ausgebaut ist die Amtliche Statistik für den industriellen Sektor. Auf Basis der bei Bund und Ländern einheitlich geführten Kartei im Produzierenden Gewerbe besteht in der Monatsberichterstattung im Bergbau und Verarbeitenden Gewerbe sowie der Erhebung der industriellen Kleinbetriebe ein aufeinander abgestimmtes Berichtssystem.

---

<sup>17</sup> Es wird hier keine vollständige Erhebung aller verfügbaren Unternehmens- oder Betriebsdaten angestrebt, sondern es soll ein kurzer Überblick über die wichtigsten Datenquellen gegeben werden. Für detailliertere Informationen sei auf die angegebene Literatur verwiesen.

Grundlage der statistischen Erhebung bei der Umsatzsteuerstatistik ist die Umsatzsteuervoranmeldung, die der Unternehmer bis zum zehnten jedes Monats bzw. des ersten Monats jedes Quartals abzugeben hat. Die Angaben werden dem Statistischen Bundesamt in zweijährigen Abständen von den zuständigen Finanzämtern zur Verfügung gestellt. Der Umsatzsteuerstatistik liegt das Unternehmenskonzept<sup>18</sup> zugrunde. Dies führt dazu, daß mehrere Betriebe eines Unternehmens als eine Einheit mit dem Jahresumsatz bei dem für den Sitz der Geschäftsleitung zuständigen Finanzamt erfaßt werden. Aus diesem Grunde ist eine regionale Zuordnung nur bedingt möglich. Dies gilt auch für die wirtschaftsbezogene Zuordnung der Unternehmen, die in der Umsatzsteuerstatistik nach dem wirtschaftlichen Schwerpunkt des Unternehmens vorgenommen wird. Aufgrund des zweijährigen Abstands der Erhebungen und der relativ späten Veröffentlichung der Ergebnisse ist eine Aktualität der Daten nicht unbedingt gegeben, d.h. auch in diesem Falle ist eine Bestandsfortschreibung notwendig, um Aufschluß über den aktuellen Unternehmensbestand zu erhalten. Ein großer Vorteil der Umsatzsteuerstatistik liegt in der weitgehenden Vollständigkeit und Zuverlässigkeit des Datenmaterials (vgl. Schumacher/Kistner 1989, S. 8ff).

Bei der Beurteilung der Amtlichen Statistik kommt Dahremöller (1988, S. 34) zu folgendem Ergebnis:

"Ist die amtliche Statistik bereits bei der Erfassung von Strukturdaten der Unternehmen bzw. Betriebe auf Stichtagsbasis unzureichend, gilt dies erst recht hinsichtlich des Ausweises des Fluktuationsgeschehens und der sich hinter den Bestandsveränderungen verbergenden dynamischen einzelwirtschaftlichen Komponenten. Über die Entwicklung identischer Unternehmen bzw. Betriebe im Zeitablauf sowie die Dynamik des ständigen Umschichtungsprozesses angebotener Arbeitsplätze können auf Basis stichtagsbezogener Daten keine Aussagen gemacht werden."

---

18 Die sekundärstatistischen Daten werden nicht zur Erstellung einer Unternehmensstatistik erhoben, an der sich die Abgrenzung des Erhebungsobjekts orientieren würde. Es lassen sich drei Grundkonzepte unterscheiden (vgl. Schumacher/Kistner 1989, S. 4):

Das Unternehmenskonzept: als Erhebungseinheit wird von rechtlich selbständigen oder wirtschaftlichen Einheiten ausgegangen (Umsatzsteuerstatistik, Berufsgenossenschaften).

Das Betriebskonzept: das Erhebungsobjekt stellt der Betrieb als örtliche Einheit und technisch-organisatorisch selbständiger Teil des Unternehmens dar (Zeitreihen des Anlagevermögens in Industriezweigen des Deutschen Instituts für Wirtschaftsforschung).

Das Betriebsstättenkonzept: als Erhebungsobjekt werden alle örtlich selbständigen Einheiten eines Unternehmens erfaßt (Beschäftigtenstatistik der Bundesanstalt für Arbeit, Arbeitsstättenzählung).

### 3.1.2 Daten der Berufsgenossenschaften

Alle Unternehmen (außer Staat und Landwirtschaft) in der Bundesrepublik Deutschland sind Zwangsmittelglied in einer der 36 gewerblichen Berufsgenossenschaften, wobei sich die Mitgliedschaft am Schwerpunkt der wirtschaftlichen Tätigkeit orientiert. Am Jahresende werden von den Berufsgenossenschaften Angaben über die Gesamtzahl der Beschäftigten einschließlich der Teilzeitkräfte, die geleisteten Arbeitsstunden und die Lohn- und Gehaltssumme der Mitgliedsunternehmen erfragt. Zudem wird die Zahl der Inhaber, die Zahl der mithelfenden Familienangehörigen und die Zahl der Arbeitstage erfaßt. Weil die Angaben der Unternehmen in regelmäßigen Abständen überprüft werden, kann von einer Vollständigkeit und Zuverlässigkeit der Daten ausgegangen werden. Eine Vergleichbarkeit der Daten mit der Amtlichen Statistik ist jedoch nicht gegeben, da die Zuständigkeit der einzelnen Berufsgenossenschaften sich in Teilbereichen nicht mit der Branchenzugehörigkeit in der Amtlichen Statistik deckt (vgl. Dahremöller 1988, S. 50ff). Die jährliche Erhebung und die Möglichkeit, einzelne Unternehmen zu identifizieren, stellen einerseits eine relativ große Aktualität sicher und ermöglichen andererseits, die Unternehmensentwicklung zumindest im Hinblick auf die erhobenen Variablen über den Zeitablauf zu verfolgen.

### 3.1.3 Die Betriebs- und Versichertendatei der Bundesanstalt für Arbeit

Mit der Einführung des einheitlichen Meldeverfahrens zur Sozialversicherung (ab 1.1.1973) wurde der Aufbau einer Beschäftigtenstatistik möglich. Erfast werden alle Arbeitnehmer und Auszubildenden, sobald sie in einem Zweig der gesetzlichen Sozialversicherung beitragspflichtig sind<sup>19</sup>. Darüber hinaus hat die Bundesanstalt für Arbeit eine Betriebsdatei aufgebaut, die über die Betriebsnummer mit der Beschäftigtenstatistik verknüpft werden kann. Die Betriebsdatei erfaßt auf Basis der Betriebsnummern den Wirtschaftszweig und die lokale Zuordnung der Betriebsstätten.

Für jeden sozialversicherungspflichtigen Beschäftigten wird unter seiner Versicherungsnummer ein Versicherungskonto geführt. Identifikationsmerkmal ist die 12-stellige Versicherungsnummer, in der auch die Betriebsnummer des Arbeitgebers enthalten ist. Bei einem Wechsel des Arbeitgebers bleibt die alte Betriebsnummer gespeichert; die Betriebsnummer des neuen Arbeitgebers wird hinzuge-

---

<sup>19</sup> Eine detaillierte Beschreibung des Meldeverfahrens findet sich bei Dahremöller 1988, S. 57f.

fügt. Zusätzlich wird der Zeitpunkt des Ausscheidens aus dem alten Betrieb und des Beginns der neuen Tätigkeit erfaßt<sup>20</sup>.

Durch die Aggregation der Datensätze der Beschäftigten mit der gleichen Betriebsnummer wurde ein neuer Datensatz erstellt, der für jeden Stichtag eine Betriebsdatei darstellt (vgl. Cramer 1987, S. 16; für die Jahre 1977-1985 war der Stichtag jeweils der 30.6.). Mit diesem Datensatz sind einerseits Querschnittsanalysen für die einzelnen Stichtage möglich, und andererseits können über eine Verknüpfung der Betriebsnummern die Unternehmen im Zeitablauf verfolgt werden. Wenn alle in der Beschäftigtenstatistik erhobenen Daten in diesen Aggregationsprozeß einbezogen werden, könnten Informationen über das Fluktuationsgeschehen, die innerbetriebliche Mobilität, die Lohnstruktur, die Ausgestaltung von Senioritätssystemen, Eintrittsarbeitsplätze und über den Grad der internen Rekrutierung gewonnen werden.

### 3.2 Unternehmensbefragungen und Panelerhebungen

Zu den wichtigsten Unternehmensbefragungen gehören die Erhebungen des Ifo-Instituts für Wirtschaftsforschung, das seit 1949 regelmäßig Befragungen durchführt.<sup>21</sup> Die Schwerpunkte der Befragungen liegen in den Bereichen: Konjunkturtest, Investitionstest, Innovationstest und Befragungen in ausgewählten Dienstleistungsbereichen. Die Ergebnisse der Umfragen werden seit 1976 in der Ifo-Datenbank gespeichert und den Teilnehmern der Befragung bzw. den Mitgliedern des Ifo-Instituts auf niedriger Aggregationsebene (z.B. auf Ebene der Produktgruppen) zur Verfügung gestellt. Andere Interessenten müssen sich mit hochaggregiertem Material (z.B. Handelsbereiche) aus den Ifo-Umfragen oder der amtlichen Statistik zufrieden geben. Die Daten des Ifo-Instituts sind wegen ihrer periodischen Erhebung als Paneldaten einzustufen. Da sie aber für die Mehrzahl der Forscher nur hochaggregiert zur Verfügung gestellt werden, sind sie allenfalls als ergänzendes Material zu den sonst verfügbaren Aggregatdaten zu betrachten.

Mit der "Bonner Stichprobe" liegt eine weitere wichtige Unternehmensstichprobe vor. Sie enthält die veröffentlichten Bilanzdaten deutscher Industrieaktiengesellschaften, insgesamt sind 717 Unternehmen seit 1960 erfaßt (Stand 1990), für die jedoch nicht immer die gesamte Zeitreihe zur Verfügung steht. Wenn auch durch das verwendete Erhebungsverfahren ein Großteil der Unternehmen und interes-

---

<sup>20</sup> Weitere Merkmale in der Versichertendatei sind: Träger der Rentenversicherung, Geburtsjahr bzw. Alter, Geschlecht, Nationalität, Wirtschaftszweig des Arbeitgebers, Arbeits- und Wohnort, ausgeübte Tätigkeit, Stellung im Beruf, Ausbildung, beitragspflichtiges Bruttoentgelt und Dauer der Berufstätigkeit.

<sup>21</sup> Für eine ausführliche Darstellung der Ifo-Umfragen siehe Oppenländer/Poser 1989 und Penzkofer 1990.

sierender Variablen (z.B. aus dem Personalbereich) nicht berücksichtigt wird, zeigt die große Anzahl von Veröffentlichungen<sup>22</sup> auf der Basis dieses Datensatzes, daß Unternehmensdaten eine zentrale Ressource für die wirtschaftswissenschaftliche Forschung darstellen (vgl. Bruse 1980, S. 128f.; für eine genauere Beschreibung der Stichprobe siehe: Geisen 1976).

Neben diesen eher allgemein gehaltenen Erhebungen sind in den letzten Jahren eine Vielzahl von Unternehmensbefragungen zu sehr unterschiedlichen Fragestellungen durchgeführt worden.<sup>23</sup> In einer groß angelegten Studie zu den "Arbeitsmarktwirkungen moderner Technologien" (vgl. Dostal 1989, Matzner/Schettkat/Wagner 1988) wurde der Versuch unternommen, mehrere Unternehmensbefragungen zusammenzuführen, was aber aufgrund der Heterogenität der hinter den einzelnen Befragungen stehenden Themengebiete mit großen Schwierigkeiten verbunden war. Von ähnlichen Problemen bei der Synopse der Ifo-Umfragen berichtet Strigel (1989, S. 334ff). Der oben bereits formulierte Mangel an Mikrodaten auf einzelwirtschaftlicher Ebene ist als Hauptursache für die vielen Unternehmensbefragungen zu sehen. Sie schließen zwar für eine bestimmte Forschergruppe die Datenlücke<sup>24</sup>, verursachen aber zum einen sehr hohe Kosten und zum anderen muß mit steigender Zahl von Befragungen von einer abnehmenden Bereitschaft zur Teilnahme seitens der Betriebe oder Unternehmen ausgegangen werden.

Der Erfolg des Sozio-ökonomischen Panels, einer Längsschnittuntersuchung, die auf Individual- und Haushaltsebene seit 1984 jährlich durchgeführt wird, hat die Diskussion um die Schaffung eines Betriebspanels als zentrale Datenressource stark belebt. Hübler (1990, S. 65f.) nennt als wesentliche Vorteile von Paneldaten:

- Erweiterung des Stichprobenumfangs und eine damit einhergehende Verbesserung der Schätzgenauigkeit;
- die Möglichkeit, alternative Verhaltenshypothesen gegeneinander zu testen, zwischen Kohorten-, Perioden- und Alterseffekten, zwischen Zeit- und Querschnitteffekten, zwischen lang- und kurzfristigen Effekten, zwischen zustandsabhängigen und genuin bedingten Heterogenitäten trennen zu können;

---

<sup>22</sup> Vgl. z.B. Albach/Burg 1984 und die dort angegebene Literatur.

<sup>23</sup> Exemplarisch für das untersuchte Spektrum seien hier nur einige genannt: Zur Empirie befristeter Arbeitsverträge: Burgbacher/Hartwig/Liebau 1989, Büchtemann 1989; zum Zusammenhang zwischen Kooperation und Produktivität: Fitzroy/Kraft 1987; zu betrieblichen Strategien zur Befriedigung von Arbeitnehmerbedürfnissen: Scholl 1986; zu den Determinanten der Arbeitsplatzentwicklung in Industriebetrieben: Fritsch 1990a; zum Rekrutierungsverhalten: Deeke/Fischer 1986; einen Überblick über Unternehmensbefragungen zu den besonderen Beschäftigungsproblemen Schwerbehinderter gibt Frick 1992.

<sup>24</sup> Die Qualität einer Befragung muß jedoch im Einzelfall beurteilt werden.

- Probleme, die sich aus Abhängigkeiten zwischen exogenen Variablen (Multikollinearität) und zwischen exogenen und unbeobachteten Einflüssen ergeben, können umgangen werden;
- Effekte, die auf wirtschaftspolitische Maßnahmen zurückgehen, werden erfaßbar.

In den USA liegt mit dem PIMS-Programm (Profit Impact of Market Strategies) ein Längsschnittdatensatz vor, der Daten über die Marktbedingungen, die Wettbewerbsposition und Indikatoren der Rentabilität und Betriebseffizienz von über 3.000 strategischen Geschäftseinheiten enthält. Dieser Datensatz ist nicht für Forschungszwecke erstellt worden, sondern es sollte Unternehmen die Möglichkeit gegeben werden, aus den "Erfahrungen anderer zu lernen". Mit Erwerb der Mitgliedschaft können die Unternehmen die Datenbank zu Auswertungszwecken nutzen. Auch wenn die Datenbank nicht zu Forschungszwecken angelegt wurde, ist sie für zahlreiche wissenschaftliche Arbeiten besonders in der Industrieökonomik genutzt worden.<sup>25</sup> Da die Teilnahme an diesem Programm freiwillig ist und in erster Linie Fragen zur strategischen Unternehmenspolitik im Mittelpunkt stehen, sind in der Datenbank überwiegend Informationen über Großunternehmen mit mehreren Geschäftseinheiten gespeichert. Es handelt sich also nicht um eine repräsentative Stichprobe; sie kann schon aufgrund des beschränkten Inhalts nicht für beliebige Forschungsfragen genutzt werden. Insbesondere für die Analyse der Personal- und Beschäftigungspolitik von Unternehmen stehen keine Informationen (außer dem gewerkschaftlichen Organisationsgrad) zur Verfügung<sup>26</sup>. Zusätzlich zur PIMS-Hauptdatenbank sind mehrere ergänzende Datenbanken entwickelt worden, die sich mit spezifischen Anwendungsbereichen befassen. Zu diesen Datenbanken zählt OASIS (Organization And Strategic Information Service), mit der Beziehungen zwischen Organisationsstruktur, Wettbewerbsstrategie und Unternehmenserfolg untersucht werden können. Obwohl die Datenbank bisher nur wenige Datensätze enthält, zeigen erste Auswertungen auf der Basis der Daten von 58 Geschäftseinheiten interessante Auswertungsmöglichkeiten. Mit einer Ausweitung der Datenbasis könnten umfangreiche Hypothesentests bezüglich des Zusammenhangs zwischen Organisationsstruktur und Unternehmenserfolg möglich werden. Inwieweit sich daraus repräsentative Aussagen

---

<sup>25</sup> Vgl. zur Konzeption und zum Inhalt der Datenbank: Buzzell/Gale 1989. Für die damit bearbeiteten Forschungsfelder vgl. ebenda, S. 247ff; Meyer/Greif 1990. Kritisch zum Konzept und den auf Basis der Daten gezogenen Schlußfolgerungen: Barzen/Wahle 1990.

Wie oben bereits erwähnt, wurde dieser Datensatz zur Analyse des Unternehmenserfolgs in Abhängigkeit von Konzentrationsgrad und Marktanteil eingesetzt. Die Ergebnisse führten zur Ablösung eines lange vorherrschenden Paradigmas in der Industrieökonomik (vgl. Buzzell/Gale 1989, S. 75ff).

<sup>26</sup> Vgl. hierzu die Fragebögen in Buzzell/Gale 1989, S. 219ff.

entwickeln lassen, hängt davon ab, ob die Datenbasis so ausgebaut werden kann, daß die erfaßten Unternehmen die Grundgesamtheit abbilden<sup>27</sup>.

In mehreren europäischen Ländern werden oder wurden bereits Betriebspanelerhebungen durchgeführt. Diese Erhebungen sind meist regional begrenzt, berücksichtigen nur einen Ausschnitt der Grundgesamtheit und haben ausgewählte Untersuchungsschwerpunkte (vgl. für Luxemburg: Allegreza 1990, für Süd-Dänemark: Hohwü 1990, für Frankreich-Nancy: Gerlach et. al. 1990). Übersicht 1 beinhaltet die wesentlichen Informationen zu den Panelerhebungen in den einzelnen Ländern.

Die Erfahrungen aus diesen Projekten zeigen, daß die Panelmortalität das Hauptproblem bei Betriebspanelerhebungen darstellt. Die lokale Begrenzung des Erhebungsraumes bzw. die geringe Größe des jeweiligen Landes oder der Region deuten auf einen anderen wichtigen Aspekt im Zusammenhang mit Betriebspanelerhebungen hin: Untersuchungen dieser Art verursachen in der Erhebungs- und Aufbereitungsphase (vor allem bei Erhebungen in größeren Regionen) erhebliche Kosten.

Zum gegenwärtigen Zeitpunkt sind in der Bundesrepublik mehrere Betriebspanel in Vorbereitung bzw. die ersten Wellen bereits durchgeführt. In dem vom Institut für Arbeitsmarkt- und Berufsforschung vorgesehenen und bereits in Planung befindlichen Betriebspanel ist für den hier interessierenden Forschungsgegenstand ein vielversprechender Ansatz zusehen. Die Initiatoren dieses Panelvorhabens begründen die Notwendigkeit eines Betriebspanels damit, daß die amtliche Statistik, die Geschäftsstatistik der Bundesanstalt für Arbeit, partielle Forschungs- und Befragungsansätze einzelner Wissenschaftsdisziplinen nur Teilaspekte des betrieblichen Beschäftigungsgeschehens erfassen, ohne aber die gesamtwirtschaftliche Nachfrageseite des Arbeitsmarktes transparent zu machen. Auch wenn sehr viele Einzelinformationen und eine Fülle von Forschungsergebnissen vorliegen, lassen sie sich selten verknüpfen (vgl. Projektgruppe Betriebspanel 1991, S. 514). Mit dem Betriebspanel des IAB wird die Schaffung einer dem Sozio-ökonomischen Panel vergleichbaren Datengrundlage für nachfrageseitige Fragestellungen angestrebt. Dabei soll den Interessen der Betriebe, der Arbeitsämter und der Forscher Rechnung getragen werden. Erste durchgeführte Tests und die erste Welle zeigen überwiegend ermutigende Ergebnisse. Möglicherweise wird aber eine dem Sozio-ökonomischen Panel vergleichbare Datenressource nicht entstehen können, weil sich in der Testphase der Fragebogen als zu umfangreich herausgestellt hat und deshalb in der im 3. Quartal 1993 durchgeführten ersten Welle ein erheblich gekürzter Fragebogen eingesetzt wurde.

---

<sup>27</sup> Vgl. zu ersten Ergebnissen: Cowherd/Luchs 1988.

Übersicht 1: Kenndaten zu den Betriebspanel-Untersuchungen in den einzelnen Ländern

	Luxemburg (altes Panel)	Luxemburg (neues Panel)	Süd-Dänemark	Nancy
I. Welle	1984	7/89-9/89	10/89-2/90	5/89-10/89
Stichprobe	geschichtet repräsentativ über alle Branchen	Vollerhebung im Verarbeitenden Gewerbe	Industriebetriebe	geschichtet kleine und mittlere Betriebe
Samplegröße	400	647	200	605
Antwortverweigerung	33% kleine 14% mittlere 18% große Betriebe	22%	30%	30%
Untersuchungsschwerpunkte	Innovation Export	Innovation Export Beschäftigung Umsatz	Innovation Export Beschäftigung kleine und mittlere Betriebe Netzwerke	Innovation Export Beschäftigung
Laufzeit	3 Jahre	unbegrenzt	5 Jahre	?
Panelmortalität	66,4% der Betriebe aus d. 1. Welle wurden in allen 3 Wellen befragt			80% der Betriebe aus der 1. Welle antworteten in der 2. Welle

Quelle: Gerlach et. al. 1990, S.19.

Mit zunehmender Breite an Informationen nimmt die Teilnahmebereitschaft der Unternehmen und Betriebe ab, so daß man sich mit einem wesentlich geringeren Maß an Informationen zufriedengeben muß und damit nicht den Interessen aller Parteien Rechnung getragen werden kann. In der ersten Welle wurden 4.356 Interviews realisiert, was einer durchschnittlichen Teilnahmequote von 71% entspricht. Bei den Fragen nach Umsatz, Lohnsumme und Beschäftigungserwartungen sind mit 39% (Umsatz), 21% (Lohnsumme) und 7% (Beschäftigungserwartungen) erhebliche Antwortverweigerungen festzustellen (vgl. Kühl 1995, S. 27). Die hohen Antwortverweigerungen bei den Fragen nach "harten" Daten zeigen die große Sensibilität der Unternehmen bei der Weitergabe solcher Informationen. Die Antwortverweigerungen dürften daher bei der Erfas-

sung von Gewinn bzw. Anlagevermögen noch höher ausfallen.<sup>28</sup> Erste Erfahrungen aus dem Pretest des Firmenpanels "Produzieren in Niedersachsen und Europäischer Binnenmarkt" weisen auf ähnliche Probleme bei der Auswahl des Befragungsverfahrens (schriftliche Befragung vs. Interview), bei der Auswahl der Befragungsperson und bei der Motivation der Unternehmen zur Teilnahme hin (vgl. Gerlach et. al. 1990, S. 21ff).

An der Universität Mannheim ist ein Betriebspanel zur Analyse der Arbeitsplatzentwicklung mit den Daten von Creditreform, der führenden Wirtschaftsauskunftsdatei, aufgebaut worden (vgl. Stahl 1991). Das Panel wird aus regelmäßigen Ziehungen von dem beim Verband der Vereine Creditreform (VVC) vorliegenden Datenbestand entwickelt. In diesem Datenbestand stellen rechtlich selbständige Unternehmen die statistische Grundeinheit dar (Unternehmenskonzept). Das Kerngeschäft von Kreditauskunftsdateien besteht in der Bereitstellung von Kreditwürdigkeitsauskünften zu einzelnen Unternehmen. Daher wurde Creditreform früher nur dann aktiv, wenn eine konkrete Anfrage bestand, was dazu führte, daß die Daten auch erst nach einer Anfrage aktualisiert wurden bzw. nur Daten mit Bedarf an Lieferantenkrediten gesammelt wurden. Daten bzw. aktualisierte Daten von Unternehmen ohne Kreditnachfrage bzw. neuer Kreditnachfrage blieben außen vor.

Stahl (1991, S. 4) führt aus, daß sich dieses Konzept in jüngerer Zeit grundsätzlich geändert habe. Er führt dies im wesentlichen auf zwei Gründe zurück: "Zum einen werden seit einigen Jahren Kreditauskünfte "auf Vorrat" online angeboten; und zum zweiten wird diese Datei in ganz anderen Zusammenhängen, z.B. für Branchenstudien und Marketingzwecke, kommerziell verwendet. Daraus ergeben sich erhebliche Konsequenzen für die wissenschaftliche Verwendbarkeit des Datenmaterials: Einerseits wird der Datensatz ständig auf dem neuesten Stand gehalten; und andererseits steigt systematisch die Zahl der erfaßten Unternehmen" (ebenda, S. 4). Derzeit werden die Informationen in der folgenden Form gesammelt: Auch ohne Vorliegen einer konkreten Anfrage werden unter anderem ständig alle öffentlichen Register und Meldungen der Tageszeitungen, die Geschäftsberichte und die veröffentlichten Bilanzen der Unternehmen ausgewertet. Zusätzlich werden in regelmäßigen Abständen (12 Monaten) die nicht beauskunfteten Unternehmen schriftlich befragt, und deren Angaben werden von Creditreform durch ein Branchenanalyseprogramm auf Konsistenz überprüft. Ergänzt werden die Daten zudem noch durch mündliche Interviews der Unternehmen und mündliche oder schriftliche Befragung von Dritten, z.B. Lieferanten, über Erfahrungen im Geschäftsverkehr mit dem beauskunfteten Unternehmen. Die Vereine

---

<sup>28</sup> Zum aktuellen Stand vgl.: Projektgruppe Betriebspanel (1994): IAB-Betriebspanel - Ergebnisse der ersten Welle 1993; Kühl 1995.

Creditreform erheben in formatierter, d.h. in maschinenlesbarer Form unter anderem folgende Daten (vgl. Stahl 1991, S. 5):

- Firmierung (Name, Anschrift, Telefon)
- Rechtsform
- Wirtschaftszweig (zwei Angaben, kompatibel zur Systematik des statistischen Bundesamtes)
- Jahresumsätze und Exportanteil
- Anzahl und Struktur der Beschäftigten
- Gründungsdatum und Umwandlungen
- Kapitalausstattung
- Mehrstufige Beteiligungsverhältnisse (Art, Anteil)
- Management (Geschäftsführung, Vorstände)
- Bilanzdaten
- Immobilien (Eigentumsverhältnisse, Wert, Belastung)
- Bankverbindungen und Kreditlinie
- Negativmerkmale (Konkurs- und Vergleichsverfahren, Scheck- und Wechselproteste)

Darüber hinaus gibt es zu diesen Bereichen Textergänzungen, die weitergehende qualitative Informationen enthalten. Für die Panelunternehmen sind ergänzend Zusatzbefragungen insbesondere zur Investitions- und Innovationstätigkeit sowie zur Inputverwendung vorgesehen.

Aus den bei Creditreform vorhandenen Datensätzen wurden zunächst 30.000 Datensätze (auf Unternehmensebene), geschichtet nach Beschäftigtengrößenklassen, gezogen. Aus diesen 30.000 Datensätzen werden dann 12.000 so ausgewählt, daß eine Kreuztabelle nach Betriebsgrößenklassen und Branchen eine hinreichend große Zellbesetzung aufweist. Um Repräsentativität hinsichtlich der Gesamtpopulation herzustellen, wurden Hochrechnungsfaktoren aus der Arbeitsstättenzählung von 1987 (aggregiert auf Unternehmensebene) ermittelt. Das so erzeugte Unternehmenspanel soll in erster Linie zur Analyse der Arbeitsplatzdynamik herangezogen werden. Insbesondere steht eine Überprüfung der von Birch (1978) aufgestellten Thesen, nach denen der größte Teil des Angebots an neuen Arbeitsplätzen in der amerikanischen Volkswirtschaft in den 70er Jahren von kleinen und mittleren Unternehmen bereitgestellt wurde, im Mittelpunkt der Analyse.

Zusätzlich zu diesen Vorhaben wird am Sonderforschungsbereich 187 der Ruhr-Universität Bochum ein Panel zu "Neuen Informationstechnologien und flexiblen Arbeitssystemen" (NIFA- Panel) als Vollerhebung bei Betrieben des deutschen Maschinenbaus mit eigener Fertigung und mindestens 20 Beschäftigten in den

alten Bundesländern durchgeführt (vgl. Flimm/Saurwein 1991)<sup>29</sup>. In jährlichem Abstand sollen bis 1996 metallverarbeitende Fertigungsbetriebe mit einem standardisierten Erhebungsinstrument schriftlich befragt werden. Auf einer relativ breiten empirischen Basis sollen repräsentative/quantitative Analysen zu vier Bereichen vorgenommen werden (vgl. Flimm/Saurwein 1991, S. 6):

- Betrieblicher Einsatz computergestützter Techniken
- Muster und Formen der betrieblichen Arbeitsorganisation
- Personalpolitik und Qualifikation
- Alternative Formen des Personaleinsatzes nach dem Prinzip der Gruppenarbeit (Schwerpunkt der ersten Welle, Querschnittsbefragung)

In einem Erfahrungsbericht zur ersten Welle des NIFA-Panels (vgl. Hauptmanns/Saurwein 1992, S. 76ff) verweisen die Autoren auf die Ausfälle bei schriftlichen Befragungen, die bei Betriebsbefragungen verstärkt auftreten. Mit dem Basisbogen des NIFA-Panels wurde letztlich ein verwertbarer Rücklauf von 47% erzielt; bei dem wesentlich umfangreicheren Hauptbogen lag die Rücklaufquote bei 30,9% (= 1.682 verwertbare Interviews). Eine Ausfalluntersuchung ergab, daß überwiegend Kleinbetriebe und Betriebe ohne Computer-Technik-Einsatz nicht an der Befragung teilnahmen. Dabei konnte in einer Ausfalluntersuchung gezeigt werden, daß nicht die Betriebsgröße die entscheidende Determinante darstellt, "... sondern die Frage, ob die Betriebe die Befragung als für sich passend und damit unter Umständen nützlich definieren. Im Falle des NIFA-Panels war dies primär die Frage, ob die befragten Betriebe computergestützte Technik einsetzen (und damit die Befragung als für sich zutreffend und unter Umständen relevant erkennen)" (Hauptmanns/Saurwein 1992, S. 83). Die durchgeführte Ausfalluntersuchung ermöglichte jedoch nicht die Entwicklung eines zufriedenstellenden Modells des Ausfallmechanismus, so daß eine Gewichtung mit dem Ziel der Anpassung an die Größen in der Grundgesamtheit zu größeren Verzerrungen geführt hätte als die Arbeit mit der ungewichteten Stichprobe. Eine Gewichtung hätte ohne die Berücksichtigung des Ausfallmechanismus zu einer Überschätzung der Betriebe mit C-Technik geführt, da Betriebe ohne C-Technik eine wesentlich höhere Ausfallquote aufweisen<sup>30</sup>.

### 3.3 Unternehmensdatenbanken

Der Markt für On-Line verfügbare Datenbanken ist kaum noch überschaubar. Allein im Bereich der Wirtschaftsdatenbanken konnte bereits im Jahr 1988 auf

<sup>29</sup> Vgl. auch: Hauptmanns 1990; Freriks 1991; Engel/Krause/Wagner 1991.

<sup>30</sup> Vgl. auch: Niggemann 1993. Für eine Analyse des Einflusses der Betriebsgröße auf den Technikeinsatz mit Daten aus der zweiten Welle des NIFA-Panels vgl. Ostendorf/Schmid 1994. Zur Dokumentation der NIFA-Datenbank siehe Rogalski 1994.

rund 1.500 Datenbanken zurückgegriffen werden, von denen 155 Datenbanken Unternehmens- und Kreditinformationen zur Verfügung stellen (vgl. Becker 1988, S. 23)<sup>31</sup>. Obwohl die Inhalte der Datenbanken in den Informationen der Hosts (Informationsanbieter) relativ gut beschrieben sind, kann ein Urteil über die Qualität der gespeicherten Daten erst nach einer Recherche in der jeweiligen Datenbank abgegeben werden. Es ist zu vermuten, daß die Unternehmen Informationen, die Einblick in die Geschäftsgeheimnisse gestatten, nicht an die Datenbankanbieter weitergeben. Die sehr kontroverse Diskussion über die Aufnahme der Jahresabschlüsse von kleinen und mittleren Kapitalgesellschaften, deren Firmendaten nur im Handelsregister hinterlegt sind, zeigt, daß man sich hier auf einem sehr sensiblen Feld bewegt (vgl. Börner 1988, Schmidt 1988)<sup>32</sup>.

Mehrere Empirische Untersuchungen zu größenspezifischen Besonderheiten der Arbeitsmarktdynamik greifen auf Kreditauskunftdateien zurück<sup>33</sup>. Am Beispiel der Arbeit von Birch (1979), wo die Daten der Dun & Breadstreet Kreditauskunftdatei genutzt werden, lassen sich die wesentlichen Probleme bei der Nutzung von Unternehmensdatenbanken verdeutlichen. Wie oben bereits erwähnt, kommt Birch für den Zeitraum 1969-1976 in den USA zu dem Ergebnis, daß fast zwei Drittel der 6,7 Millionen neu geschaffenen Arbeitsplätze auf Unternehmen mit bis zu 20 Beschäftigten entfallen. Die Ergebnisse von Birch mußten erheblich nach unten korrigiert werden (vgl. Dahremöller 1988, S. 14ff). Dafür werden folgende Gründe angeführt:

- Unternehmenszugänge und -abgänge werden nur mit erheblicher Verzögerung erfaßt. In die Datei wurden in erster Linie stark expandierende Unternehmen aufgenommen, da diese in größerem Ausmaß Kredite nachfragen.
- Eine Trennung von rechtlich selbständigen Unternehmen und zugehörigen Unternehmensteilen ist nicht immer möglich, so daß Doppelzählungen die Folge sind.
- Kleine und mittlere Unternehmen sind in der Datenbasis stark unterrepräsentiert.
- Ein großer Teil der erfaßten Datensätze der Dun & Breadstreet Kreditauskunftdatei waren nicht aktualisiert. In diesen Fällen wurden die Daten des Vorjahres einfach in das Folgejahr übertragen.

Neben extrem hohen Kosten (vgl. z.B.: Dialog Database Catalog 1989) verhindert die fehlende Kenntnis der Qualität der gespeicherten Daten und die mögli-

---

<sup>31</sup> Vgl. auch: Foster 1984, S. 86ff; Foster 1984a, S. 174ff.; Monk/Landis/Monk 1988, S. 86-532; Tudor 1989, S. 47-365.

<sup>32</sup> Bei den im Handelsregister hinterlegten Daten über Unternehmen handelt es sich um für Forschungszwecke kaum brauchbare Daten.

<sup>33</sup> Für eine Übersicht vgl. Fritsch/Hull 1987, S. 154f.

cherweise mangelnde Repräsentativität eine Nutzung dieser Datenquellen zur Schaffung einer Unternehmensstichprobe. Aufschluß über die Qualität der Daten würde wohl nur eine Überprüfung in jedem Einzelfall geben.

### 3.4 Fallstudien

Dieser Typ von Daten ermöglicht den am weitesten gehenden Einblick in potentielle Wirkungszusammenhänge auf einzelwirtschaftlicher Ebene. Die wohl bedeutendste Fallstudie für die Bundesrepublik liegt für einen großen Maschinenbaubetrieb ("Südwerk") vor; die Daten wurden aus den Personaldateien des Unternehmens gewonnen. Zum einen werden die Grunddaten für alle Beschäftigten erfaßt, und zum anderen können die Daten aus dem laufenden Arbeitsprozeß mit in die Analysen einbezogen werden. Die Daten stehen für den Zeitraum von 1976-1984 zur Verfügung, so daß Längsschnittbetrachtungen über neun Jahre möglich sind (vgl. Köhler/Schultz-Wild 1989). Die seltene Gelegenheit des Zugriffs auf prozeßproduzierte und anonymisierte betriebliche Personaldaten eröffnet ein breites Feld von Analysemöglichkeiten zur Überprüfung zentraler Fragen der arbeitsmarkttheoretischen, industriesoziologischen und organisationswissenschaftlichen Forschung (vgl. Köhler/Preisendörfer 1989, S. 7). Brüderl (1989, S. 83) kann auf dieser Datenbasis einen konkaven Verlauf der Einkommen in Abhängigkeit von der Betriebszugehörigkeitsdauer nachweisen. Der hohe Einfluß der Variable "Aufstieg in der Jobhierarchie" auf den Steigungsparameter des Lohnprofils läßt auf eine starke Kopplung der Löhne an Arbeitsplätze und eine hierarchische Lohnstruktur schließen (vgl. ebenda, S. 85). Eine Analyse der Eintrittskohorten von 1976/77 zeigt eine mit der Betriebszugehörigkeitsdauer zunehmende durchschnittliche Lohngruppenentwicklung, die jedoch in Abhängigkeit von Geschlecht und Nationalität im Niveau variiert (vgl. Grüner 1989, S. 111f.). Die Verläufe der innerbetrieblichen Mobilitätspfade sind stark abhängig davon, ob ein Arbeiter im Fertigungs- oder im Servicebereich eingestellt wird. Bei einer Einstellung im Servicebereich ist eine höhere Hierarchiestufe leichter zu erreichen als bei einer Einstellung im Fertigungsbereich.

Weitere Beispiele für die "Tiefe" der durch Fallstudien zu erhaltenden Informationen sind die Fallstudien von Windolf/Hohn (1984) und Dombois (1982), die Einblicke in den Einfluß der betrieblichen Normenstruktur auf den betrieblichen Leistungserstellungsprozeß ermöglichen. Fallstudien zeichnen sich zwar durch sehr detaillierte Informationen über Mikroeinheiten aus, die Möglichkeit der Verallgemeinerung der Ergebnisse ist aber infolge der meist geringen Fallzahl selten gegeben. Gerlach et. al. (1990, S. 10) sehen den Idealfall in einer Vielzahl von Fallstudien, bei denen quantitative Daten erfaßt und die ständig wiederholt werden.

### **III. Konstruktion synthetischer Daten mittels Simulated Annealing: Theoretische Analyse des Verfahrens und Experimente**

Die exemplarische Darstellung der einzelnen Datenquellen und -typen zeigt, daß es einen "Königsweg" zur Lösung des Datenproblems nicht gibt. Es konnte jedoch herausgearbeitet werden, daß durch vorhandene Mikrodaten über Betriebe oder Unternehmen ein erheblicher Erkenntnisfortschritt zu erwarten wäre, weil mikroökonomisch abgeleitete Theorien durch Daten auf gleicher Ebene testbar würden. Hinreichend disaggregierte Daten (räumlich und zeitlich) gestatten zudem jede beliebige Aggregation, so daß sowohl Aussagen über die Mikroeinheiten als auch über die Aggregate möglich sind.

Mit der Arbeitsstättenzählung von 1987 liegt erstmals seit 1970 eine Totalerhebung der bundesrepublikanischen Unternehmen und Arbeitsstätten (alte Bundesländer) vor, die als Ausgangspunkt für die Konstruktion einer synthetischen Unternehmensstichprobe genutzt wird. Da bei dieser Erhebung Auskunftspflicht bestand, kann von einer hohen Genauigkeit des Datenmaterials ausgegangen werden. Zu den veröffentlichten Aggregatinformationen gehört eine empirische Verteilung, deren Eigenschaften bis auf diese Aggregatinformationen unbekannt sind. Es ist leicht einsehbar, daß zu einer Randverteilung beliebig viele Verteilungen denkbar sind. Gesucht ist also eine beste Schätzung für die unbekannt verteilung in der Grundgesamtheit. Zum einen ist es sicherlich kein triviales Problem, eine mit vielen Randinformationen möglichst gut übereinstimmende Stichprobe zu ermitteln, und zum anderen ist nicht zu überprüfen, ob eine "einfach ausgezählte" Stichprobe, die mit den Randverteilungen verträglich ist, als Schätzung für die unbekannt verteilung gelten kann.

Im folgenden wird daher ein Verfahren zur Konstruktion einer synthetischen Stichprobe vorgestellt, das die vorgegebenen Randverteilungen gut approximiert, also einen "absolut kostenminimalen Zustand" erreicht und zudem "asymptotisch entropiemaximale" Verteilungen erzeugt. Das Prinzip der "maximalen Entropie" wird in der Literatur als die theoretische Leitvorstellung für die Konstruktion synthetischer Mikrodaten überhaupt bezeichnet (vgl. Gyárfás 1990, S. 56).

#### **1. Die Grundidee bei der Konstruktion synthetischer Mikrodaten mittels Simulated Annealing**

Ausgehend von aggregierten Statistiken (Randverteilungen) verschiedener Merkmale einer unbekannt Grundgesamtheit sollen fiktive, aber mit diesen Randverteilungen verträgliche Stichprobenelemente rekonstruiert werden. Zu diesem Zweck kommt ein Algorithmus, das "Simulated Annealing", zur Anwendung, bei dem, ausgehend von einer beliebigen Anfangsverteilung für die zu generierenden Variablen, die Merkmalsausprägungen der Variablen so verändert

werden, daß eine möglichst gute Reproduktion der zugrundegelegten Randverteilungen erfolgt.

In jedem Iterationsschritt werden die Merkmalsausprägungen eines zufällig ausgewählten Falles verändert, und die so neu konstruierte Stichprobe wird mit den Randverteilungen verglichen. Reproduziert die neue Stichprobe die Randverteilungen besser als die vorherige, wird die Veränderung akzeptiert und die neue Stichprobe als Grundlage für weitere Iterationen benutzt. Dieser iterative Prozeß wird zum einen durch eine Kostenfunktion, die die Abweichung von den Randverteilungen mißt, und zum anderen durch einen "Kühlplan", der die Zahl der akzeptierten Verschlechterungen sukzessive absenkt, gesteuert. Damit der Prozeß nicht in einem lokalen Minimum der Kostenfunktion hängenbleibt, müssen Verschlechterungen (im Sinne von einem größeren Wert der Kostenfunktion) während des Iterationsprozesses akzeptiert werden.

Die durch dieses Verfahren erzeugte Stichprobe weicht nur soweit von der Ausgangsverteilung ab, wie es die zu approximierenden Randverteilungen erfordern. Ist die Ausgangsverteilung eine Gleichverteilung, dann ist die resultierende Verteilung die "entropiemaximale Verteilung" bei Erfüllung der vorgegebenen Randbedingungen.

## **2. Einordnung synthetischer Daten in den Beurteilungsrahmen**

Im Vordergrund der vorliegenden Arbeit steht die Frage, inwieweit der hier vorgeschlagene Weg - Konstruktion einer synthetischen Unternehmensstichprobe - zu einer "sinnvollen" Reduzierung des Datenproblems auf einzelwirtschaftlicher Ebene führen kann. Dabei erfolgt die Auswahl der zu generierenden Variablen vor dem Hintergrund einer substanzwissenschaftlichen Fragestellung (Determinanten der betrieblichen Schwerbehindertenbeschäftigung, vgl. Abschnitt IV.1 u. IV.2). Zunächst wird der gewählte Lösungsansatz den in Abschnitt II.1 formulierten Kriterien gegenübergestellt.

### **Meß- und Spezifikationsfehler**

Ausgangspunkt für die Konstruktion der synthetischen Stichprobe bilden vom Statistischen Bundesamt (Arbeitsstättenzählung 1987), vom Deutschen Institut für Wirtschaftsforschung und von der Bundesanstalt für Arbeit (Auswertung der Arbeitgeberanzeigen nach § 13 Abs. 2 SchwbG) veröffentlichte Aggregatdaten, die durch das unten dargestellte Verfahren disaggregiert und miteinander verknüpft werden. Aufgrund der Teilnahmeverpflichtung und vorhandener Sanktionsmöglichkeiten bei Auskunftsverweigerung oder Falschangaben dürften Meß-

fehler eine untergeordnete Rolle spielen<sup>34</sup>. Da die Auswahl der erhobenen bzw. veröffentlichten Daten nicht beeinflusst werden kann, sind Operationalisierungen notwendig, die die theoretischen Größen möglichst gut abbilden; die Beurteilung einer adäquaten Spezifikation muß wohl für jeden Einzelfall erfolgen. Das Operationalisierungsproblem stellt sich jedoch auch bei der Nutzung anderer Datenquellen: Bei jeder sekundär-statistischen Auswertung ist man auf die erhobenen Größen festgelegt; es können allenfalls unterschiedliche Datensätze miteinander kombiniert werden.<sup>35</sup> Selbst bei originär für eine konkrete Forschungsfrage erhobenen Datensätzen stellt sich das Problem der Operationalisierung derart, daß sich viele theoretische Größen (z.B. Humankapital, Mobilität etc.) nicht direkt erheben lassen (vgl. Friedrichs 1985, S. 77f).

### Repräsentativität

Wenn die auf Basis der synthetischen Stichprobe berechneten Randverteilungen sich nicht wesentlich von den vorgegebenen Randverteilungen unterscheiden, dann kann die Stichprobe als repräsentativ bezüglich der in den Randverteilungen abgebildeten Merkmale angesehen werden. In dieser Anwendung wird eine Stichprobe erzeugt, die für Unternehmen und Beschäftigte repräsentativ in Bezug auf Beschäftigtengrößenklassen und Wirtschaftszweige ist. Auch bei originär erhobenen Stichproben wird die Repräsentativität aufgrund der Abbildung bestimmter, als wichtig erachteter Randverteilungen beurteilt. Oft wird Repräsentativität von nach der Erhebungsphase nicht repräsentativen Stichproben durch Berechnung von Gewichtungsfaktoren hergestellt, die auf Basis der auch bei der Konstruktion synthetischer Stichproben verwendeten Aggregatdaten berechnet werden<sup>36</sup>.

Am Kriterium der Repräsentativität läßt sich ein besonderes Problem von Unternehmensdaten verdeutlichen. Wichtigste Voraussetzung zur Beurteilung der Repräsentativität ist die Kenntnis der Grundgesamtheit. Frick (1992, S. 155) zeichnet ein sehr negatives Bild: "Ein weiteres Problem stellt ... die unzureichende Kenntnis von Umfang und Struktur der Grundgesamtheit, d.h. aller Betriebe, die Personal beschäftigen, dar. Eine aktuelle und zugängliche Liste dieser Betriebe

---

<sup>34</sup> Fehler oder Abweichungen können jedoch bei einem Wechsel der Analyseebene (Arbeitsstätte oder Unternehmen) auftreten (vgl. die Ausführungen unten zum Aggregationsniveau).

<sup>35</sup> Zu den Problemen bei der Verknüpfung verschiedener Unternehmensbefragungen vgl. z.B.: Dostal 1986, Matzner/Schettkat/Wagner 1988, Strigel 1989, S. 334ff).

<sup>36</sup> Frick (1992, S. 162) beurteilt die Repräsentativität (nach Branchen und Größenklassen) einer aus den Arbeitgeberanzeigen gezogenen Stichprobe durch den Vergleich mit den jeweils eindimensionalen Randverteilungen (Branche und Größenklasse).

existiert nicht; die bei den Betriebsnummernstellen der Arbeitsämter geführten Karteien unterliegen dem Datenschutz und gelten unter Experten als "verschmutzt"."

Da mit der Arbeitsstättenzählung von 1987 zum ersten mal seit 1970 eine Vollerhebung der bundesrepublikanischen Unternehmen und Arbeitsstätten (alte Bundesländer) vorliegt, wird diese als Grundlage für die Datenkonstruktion gewählt. Grundsätzlich sind jedoch viele unterschiedliche Aggregatinformationen zur Konstruktion einer synthetischen Stichprobe geeignet. Eine detaillierte Beurteilung zur Repräsentativität der synthetischen Stichprobe erfolgt in Kapitel IV.

### Strukturelle Veränderungen

Mehrere empirische Studien zu der Entwicklung von Unternehmen im Zeitablauf (Gründungsrate, Schließungsrate) und deren Einfluß auf die Beschäftigungsentwicklung zeigen eine erhebliche Dynamik in der Grundgesamtheit. In einer Studie zu den "Überlebenschancen und Beschäftigungswirkungen neugegründeter Industriebetriebe" errechnet Schasse (1992, S. 35) eine durchschnittliche Neugründungsrate von 3,3%<sup>37</sup> jährlich (Industriebetriebe mit weniger als 50 tätigen Personen in Niedersachsen) im Zeitraum 1979-1990. Von den neugegründeten Unternehmen existieren nach einem Jahr 17%, nach fünf Jahren knapp 36% und nach 10 Jahren ca. 50% nicht mehr (vgl. Schasse 1992, S. 39ff).

Gründungs- und Schließungsrate variieren in Abhängigkeit von der Betriebsgrößenklasse, Branche und der Region. Wesentlich höher wird die Gründungs- und Schließungsrate von Cramer (1986, S. 17, S. 24ff) ausgewiesen. Er ermittelt auf Basis der Beschäftigtenstatistik sogar eine jährliche Gründungsrate von etwa 11,7% für die Jahre 1978-1985 und eine durchschnittliche Schließungsrate von 10,4% (siehe Tabelle 2, vgl. auch Frick 1992, S. 77).<sup>38</sup>

Obwohl der Saldo zwischen Neugründungen und Schließungen sowohl für die Zahl der Betriebe als auch für die Zahl der Beschäftigten sehr klein ist, verbirgt sich hinter diesem Saldo eine erhebliche Dynamik, die sich durch eine einmalige Querschnittsbefragung nicht abbilden läßt. Sofern sich die Struktur der Variablenbeziehungen im Zeitablauf nicht ändert (erste Dimension des Problems, vgl. Abschnitt II.1), ist nur eine quantitative Variation der  $\hat{b}_{it}$  zu erwarten. Bewirken die oben skizzierten dynamischen Prozesse (unter vielen anderen denkbaren

---

<sup>37</sup> Die Neugründungsrate wird gemessen als Anteil der neugegründeten Betriebe an allen bestehenden Betrieben.

<sup>38</sup> Zur Entwicklung von Gründungs- und Schließungsrate vgl. auch Gerlach/Wagner 1992, S. 7f.. Die von Cramer berechneten Werte sind sicherlich zu hoch, da in den Zahlen zu Neugründungen und Schließungen auch lediglich die Vergabe neuer Betriebsnummern enthalten ist.

strukturellen Veränderungen) jedoch eine veränderte Struktur der Variablenbeziehungen oder werden neue Einflußfaktoren im Zeitablauf relevant, dann ist eine Aussage über den Erhebungszeitpunkt hinaus nicht möglich.<sup>39</sup>

Tabelle 2: Schließungs- und Gründungsraten von 1978-1985 und deren Einfluß auf den Beschäftigtenstand (alle Wirtschaftszweige, ohne Landwirtschaft, Post, Bahn, Gebietskörperschaften und Sozialversicherung 1978-1985)\*

Jahr	Gründungen		Schließungen		Veränderung	
	Betriebe	Beschäftigte	Betriebe	Beschäftigte	Betriebe	Beschäftigte
1978	12,3	3,8	10,7	2,9	1,6	0,7
1979	11,8	2,8	9,6	2,0	2,2	0,8
1980	11,8	2,7	10,0	2,0	1,8	0,8
1981	11,5	2,8	10,5	2,1	1,0	0,7
1982	11,8	2,7	11,4	2,6	0,4	0,2
1983	10,7	2,4	10,0	2,1	0,7	0,2
1984	11,8	2,9	10,5	2,4	1,3	0,5
1985	11,6	2,7	10,8	2,3	0,8	0,4
Ø	11,7	2,85	10,4	2,3	1,2	0,5

\* Alle Angaben in % gemessen an der Gesamtzahl der Unternehmen bzw. Beschäftigten

Quelle: Cramer 1986, S. 24f., eigene Berechnungen.

Wenn trotz dieser Probleme die Arbeitsstättenzählung von 1987 als Grundlage für die Datenkonstruktion gewählt wird, dann geschieht dies aus folgenden Gründen:

Zum einen ist diese Erhebung die genaueste Erhebung der Grundgesamtheit, auf die zurückgegriffen werden kann, das heißt Arbeitsstätten und Unternehmen sind genau abgegrenzt, wohingegen eine Nutzung der Beschäftigtenstatistik einen Teil der Beschäftigten und Unternehmen nicht berücksichtigen würde und darüber hinaus keine Aussage darüber möglich ist, ob die veröffentlichten Daten auf Arbeitsstätten-, Betriebs- oder Unternehmensebene basieren; die Betriebsnummern-datei gilt als "verschmutzt".

Zum anderen soll die generelle Eignung des Verfahrens vor dem Hintergrund einer substanzwissenschaftlichen Fragestellung überprüft werden, indem die Er-

<sup>39</sup> Diese Probleme entstehen auch bei originär erhobenen Querschnittsdaten, beispielsweise bei einmalig durchgeführten Unternehmensbefragungen.

gebnisse von Frick (1992) zu den Determinanten der Schwerbehindertenbeschäftigung, die auf Basis von Branchendaten für den Zeitraum 1982-1986 gewonnen wurden<sup>40</sup>, mit den auf Unternehmensdaten basierenden Ergebnissen verglichen werden. Da die Beobachtungszeiträume nicht weit auseinander liegen, können Strukturveränderungen vernachlässigt werden. Zudem läßt sich überprüfen, ob sich durch die Wahl einer anderen Datengrundlage die Ergebnisse verändern.

Um eine für den heutigen Zeitpunkt geltende Struktur durch eine synthetische Stichprobe abzubilden, könnte eine fortgeschriebene Arbeitsstättenzählung oder anderes aktuelles Datenmaterial verwendet werden. Die größten Schwierigkeiten bei der Verwendung von aktuellem und zuverlässigem Datenmaterial sind in der sehr verzögerten Bereitstellung dieser Daten zu sehen (vgl. z.B. Dahremöller 1988).<sup>41</sup>

### **Aggregationsniveau**

Die synthetische Stichprobe erfüllt sicherlich nicht das Kriterium einer adäquaten zeitlichen Aggregation. Da aber eine erste Beurteilung synthetischer Daten im Vordergrund steht und in weiteren Schritten mehrere Perioden hinzusimuliert werden könnten, sofern die Veränderungen von einer Periode zur nächsten bekannt sind, wird diese Unzulänglichkeit hier in Kauf genommen.<sup>42</sup>

Wie oben bereits ausgeführt, sind einzelwirtschaftliche Einheiten Gegenstand von Analysen der betrieblichen Arbeitskräftenachfrage, und die Bereitstellung entsprechender Daten steht im Mittelpunkt dieser Arbeit. Als Erhebungseinheit in der Arbeitsstättenzählung 1987 wurde die Arbeitsstätte gewählt. "Als Arbeitsstätte gilt jede örtliche Einheit (i. Orig. hervorgehoben) (ein Grundstück oder abgegrenzte Räumlichkeit), in der eine Person oder mehrere Personen unter einheitlicher Leitung regelmäßig haupt- oder nebenberuflich erwerbstätig sind" (Statistisches Bundesamt 1989, Fachserie 2, Heft 1, S. 10). Von den einzelnen Niederlassungen werden Angaben für das gesamte Unternehmen erfragt, auf deren Basis die Ergebnisse für die Unternehmensebene ermittelt werden (vgl. ebenda, S. 10).

---

<sup>40</sup> Frick (1992) interpretiert die Daten nicht im Längsschnitt, sondern faßt die einzelnen Jahresdaten zu einem Querschnittsdatensatz zusammen.

<sup>41</sup> Die ersten Ergebnisse der Arbeitsstättenzählung von 1987 lagen erst ab Mitte 1989 vor (vgl. Statistisches Bundesamt 1989). Gegeben die oben erwähnte Dynamik im Gründungs- und/oder Schließungsprozeß von Unternehmen ist es denkbar, daß es zu nennenswerten Veränderungen in der Grundgesamtheit bis zum Zeitpunkt der Veröffentlichung der Daten gekommen ist.

Die Ergebnisse werden sowohl auf Arbeitsstättenenebene (örtliche Einheiten) als auch auf Unternehmensebene (wirtschaftliche Einheit) veröffentlicht. "Als Unternehmen gilt die kleinste rechtlich selbständige Einheit, die aus handels- und/oder steuerrechtlichen Gründen Bücher führt und Ertrag ermittelt" (ebenda, S. 11). In der überwiegenden Zahl der Fälle sind Unternehmen und Arbeitsstätte identisch, d.h. das Unternehmen besteht nur aus einer einzigen Arbeitsstätte. Von den 2.097.853 Unternehmen in den acht Hauptabteilungen (ohne "Organisationen ohne Erwerbscharakter" und "Gebietskörperschaften und Sozialversicherung") hatten 1.998.015 Unternehmen nur eine Arbeitsstätte (95,2%). Bei den Beschäftigten ist die Situation dagegen völlig anders: nur 48,9% der Beschäftigten sind in Unternehmen mit einer Arbeitsstätte beschäftigt.

Die Wahl der Abbildungseinheit hat wesentlichen Einfluß auf die quantitative Ausprägung der einzelnen Größen in den Wirtschaftszweigen. Legt man das Unternehmenskonzept zugrunde, dann wird beispielsweise die Zahl der Beschäftigten auf der Ebene der Hauptabteilungen zum Teil höher ausgewiesen als auf Basis des Arbeitsstättenkonzepts. Für die Hauptabteilung 1 (Energie- und Wasserversorgung, Bergbau) werden auf Arbeitsstättenenebene 401.584 Beschäftigte angegeben und auf Unternehmensebene 485.183 (vgl. Statistisches Bundesamt 1990, Fachserie 2, Heft 11, S. 12 u. 96). Diese Abweichung ist auf die gewählte Zuordnung zu den Wirtschaftszweigen zurückzuführen. Arbeitsstätten bzw. Unternehmen werden dem Wirtschaftszweig zugeordnet, in dem der Schwerpunkt ihrer Tätigkeit liegt. Es ist leicht einsehbar, daß bei Mehrbetriebsunternehmenein Teil der Arbeitsstätten in anderen Wirtschaftszweigen tätig ist als das Unternehmen selbst, Arbeitsstätten des Unternehmens damit abweichend zum Unternehmen eingruppiert werden. Es läßt sich zeigen, daß mit einer detaillierteren Gliederung der Wirtschaftszweige die Abweichungen in Abhängigkeit von der Darstellungseinheit zunehmen.

Die Auswahl der Analyseeinheit erfolgt unter rein pragmatischen Gesichtspunkten: Es werden die Aggregatdaten auf Unternehmensebene aus der Arbeitsstättenzählung verwendet, zum einen, weil in den meisten Fällen Unternehmen und Arbeitsstätte zusammenfallen, und zum anderen, weil eine Vergleichbarkeit mit anderen Statistiken dann leichter möglich ist.<sup>43</sup> Die Informationen zur Berechnung der Schwerbehindertenquoten basieren jedoch auf einem Betriebskonzept. Für das Jahr 1987 werden auf Basis der Arbeitgeberanzeigen 179.080 Betriebe/Dienststellen mit 16 und mehr Beschäftigten ausgewiesen (vgl. Amtliche

42 Wenn man davon ausgeht, daß eine Vielzahl empirischer Studien auf Aggregatdatenaufbau, dann ist eine Analyse auf Basis von Unternehmens- oder Betriebsdaten, die zudem noch zu geringen Kosten bereitgestellt werden können, als Fortschritt zu betrachten, wenn die synthetische Stichprobe die Struktur der Grundgesamtheit approximativ abbildet.

43 Zu den unterschiedlichen Konzepten vgl. z.B.: Schumacher/Kistner 1989, S. 4f.

Nachrichten der Bundesanstalt für Arbeit 1988, S. 1755). Werden den Simulationen Daten auf Arbeitsstättenbasis zugrundegelegt, dann wird die Zahl der beschäftigungspflichtigen Betriebe stark überschätzt, obwohl die Zahl der Arbeitsstätten nach Branchen und Beschäftigtengrößenklassen annähernd exakt abgebildet wird. Eine zu große Zahl beschäftigungspflichtiger Betriebe bedingt folglich auch eine Überschätzung der Zahl der beschäftigten Schwerbehinderten. Bei Verwendung der Informationen auf Unternehmensebene wird zwar die Zahl der beschäftigungspflichtigen Unternehmen zu klein, die beschäftigten Schwerbehinderten werden jedoch recht gut abgebildet (vgl. Abschnitt IV.4.3.1, Tabelle 29).

Wenn man eine einzelwirtschaftliche Perspektive voraussetzt, dann kann die Auswahl der Analyseebene "Unternehmen" als angemessen betrachtet werden, da empirische Tests über das Verhalten einzelwirtschaftlicher Einheiten durchgeführt werden können. Zudem ist davon auszugehen, daß wichtige unternehmerische Entscheidungen wie die Ausgestaltung der Personalpolitik auf Unternehmensebene getroffen werden und nicht auf Betriebs- und Arbeitsstättenebene. Als Unternehmen wird beispielsweise nicht der Konzern erfaßt, sondern wie oben beschrieben die kleinste rechtlich selbständige Einheit, die aus handels- und/oder steuerrechtlichen Gründen Bücher führt und Ertrag ermittelt.

Zusammenfassend kann festgestellt werden, daß die hier gewählte Vorgehensweise zur Reduzierung des Datenproblems für einzelwirtschaftliche Einheiten zwar keinen Idealtypus darstellt, aber eine kostengünstige Alternative zur Bereitstellung von Unternehmensdaten ist, sofern eine gute Approximation an die Grundgesamtheit erreicht wird. Wenn auch zunächst keine Längsschnittbetrachtungen möglich sind, sind die in Abschnitt II.1 formulierten Kriterien weitgehend erfüllt oder nicht mehr als bei anderen Vorgehensweisen verletzt.

### 3. Theoretische Analyse der Konvergenzeigenschaften des Simulated Annealing

#### 3.1 Definition des Optimierungsproblems<sup>44</sup>

Die "Fälle"<sup>45</sup> der zu generierenden Unternehmensstichprobe seien durch einen Merkmalsvektor  $v = (v_1, \dots, v_z)'$  beschrieben<sup>46</sup>, wobei es sich nur um diskrete Merkmale handeln soll. Die gemeinsame Verteilung aller Merkmalskombinationen

---

<sup>44</sup> Vgl. zu diesem Abschnitt: Graef/Blien 1989, S. 123ff; Gyárfás 1988, S. 15f.; Gyárfás/Quinke 1991, S. 1ff; Gyárfás/Quinke 1993; Merz 1983, S. 344ff.

<sup>45</sup> Mit "Fall" oder Beobachtungseinheit (Merkmalsträger) ist beispielsweise ein Unternehmen oder ein Individuum gemeint.

<sup>46</sup>  $v'$  bezeichnet den transponierten Vektor.

nen ist dann eine Tabelle mit  $k$  Zellen. Gesucht ist die unbekannte Verteilung  $p^0 = (p_1^0, \dots, p_k^0)'$  mit

$$(5) \quad p^0 \in P := \left\{ p \mid p_i \geq 0, \sum_{i=1}^k p_i = 1 \right\}.$$

Eine Verteilung  $p \in P$  wird als gemeinsame Verteilung der Merkmalskombinationen der beteiligten Variablen interpretiert. Zur Abbildung oder Schätzung der Verteilung  $p^0$  gibt es mehrere Möglichkeiten: Erheben einer originären Stichprobe, Interpretation von Aggregatdaten als Mikrodaten (vgl. z.B. Frick 1992, S. 167)<sup>47</sup> oder wie im vorliegenden Fall durch eine synthetische Stichprobe.<sup>48</sup>

Es wird davon ausgegangen, daß keine Mikrodaten zur Verfügung stehen und nur aggregierte Informationen  $r = (r_1, \dots, r_m)'$  der Verteilung  $p^0$  existieren. Zudem seien Informationen über die Verteilung  $p^0$  in Form einer a priori-Verteilung  $q = (q_1, \dots, q_k)$  gegeben. Die Verknüpfung der Verteilung  $p^0$  und den aggregierten Informationen ist durch die Beziehung

$$(6) \quad Bp^0 = r$$

gegeben. Die Elemente der Matrix  $B$  sind von der Gestalt der aggregierten Informationen abhängig: Soll an relative Häufigkeiten angepaßt werden, dann enthält die Matrix  $B$  nur Nullen und Einsen, im Falle von Mittelwerten sind die Elemente von  $B$  die Variablenwerte dividiert durch die Gesamtzahl der Fälle. Wenn an gewichtete Summen von Variablenwerten angepaßt werden soll, dann stehen in der betreffenden Zeile von  $B$  die Variablenwerte.<sup>49</sup>

Für die Verteilung  $q$  wird in den meisten Fällen dem "Prinzip vom unzureichenden Grunde" folgend eine Gleichverteilung unterstellt. In der hier vorgestellten Anwendung ist jedoch für verschiedene Größen die empirische Verteilung bekannt (z.B.: Zahl der Betriebe nach Beschäftigtengrößenklassen und Branchen; Durchschnittslohn pro Beschäftigten nach Beschäftigtengrößenklassen und Branchen). Da die Variable Beschäftigte metrisch skaliert ist<sup>50</sup>, kann die entsprechende

<sup>47</sup> Frick verwendet zur Begründung seiner Vorgehensweise das Konstrukt der "repräsentativen Firma" (vgl. Marshall 1961, S. 317-318), bzw. des "durchschnittlichen Betriebes" (vgl. Stinchcombe 1965, S. 143).

<sup>48</sup> Die Abbildungsgenauigkeit der einzelnen Ansätze hängt im wesentlichen davon ab, inwieweit die Kriterien aus Abschnitt II.1 erfüllt sind.

<sup>49</sup> Vgl. Wauschkuhn 1982, S. 18ff. Der Autor nennt weitere mögliche lineare Nebenbedingungen und erläutert an einem Beispiel die Gestalt der Matrix  $B$  in Abhängigkeit von verschiedenen Nebenbedingungen.

<sup>50</sup> Zu den verschiedenen Skalenniveaus vgl. z.B.: Bley Müller/Gehlert/Gülicher 1985, S.3.

empirische Verteilung zur Erzeugung entsprechend verteilter Zufallszahlen herangezogen werden.<sup>51</sup> Unterstellt man für  $q$  die Gleichverteilung bzw. eine aus den empirischen Verteilungen abgeleitete Verteilung, dann wird  $q$  die Bedingung (6) nicht erfüllen. Die Menge  $D_0$  sei definiert als die Menge aller Verteilungen aus  $P$ , die die Bedingung (6) erfüllen:

$$(7) \quad D_0 := \{p \in P \mid Bp = r\} \neq \emptyset. \text{ } ^{52}$$

Bei einer gegebenen Zielfunktion und  $\text{Rang}(B) = m$  und  $\text{Rang}(B, r) = m$  ist das Gleichungssystem lösbar und das Problem einer direkten Optimierung zugänglich (vgl. Wauschkuhn 1982, S. 22).

Je nach gewählter Zielfunktion<sup>53</sup> handelt es sich um ein nicht-lineares Optimierungsproblem, das mit Hilfe iterativer Verfahren gelöst werden muß. Da in dieser Anwendung metrische Größen mit theoretisch unbegrenzt vielen Merkmalsausprägungen abgebildet werden, ist an eine große Zahl von Zellen anzupassen. Die Zahl der anzupassenden Zellen wächst mit der Zahl der Merkmalsausprägungen multiplikativ, so daß Verfahren der direkten Optimierung nicht einsetzbar sind ("Kombinatorische Explosion"); es handelt sich folglich um ein kombinatorisches Optimierungsproblem.

Zur Lösung derartiger Probleme eignet sich ein Algorithmus, das "Simulated Annealing", weil es nicht auf Basis aller Zellen, sondern auf Basis der synthetischen Stichprobe arbeitet (vgl. Paaß 1988, S. 1). Es ist zu prüfen, ob das gewählte Verfahren in der Lage ist, Verteilungen aus der Menge  $D_0$  zu erzeugen (Konvergenz zum globalen Minimum)<sup>54</sup>, und welche Zielfunktion die beste Schätzung  $\hat{p}$  für die wahre Verteilung  $p^0$  aus der Menge  $D_0$  sicherstellt. Es ist durchaus denkbar, daß Bedingung (6) durch eine große Zahl von Verteilungen erfüllt wird, so daß die Auswahl der Zielfunktion die synthetisch erzeugte Verteilung erst eindeutig determiniert.

51 Vgl. dazu ausführlich Abschnitt IV.4.2.

52 Diese Gleichung gilt nur, wenn die Nebenbedingungen widerspruchsfrei sind. In der hier vorliegenden Anwendung ist dies jedoch vermutlich nicht der Fall, da Aggregatinformationen aus mehreren Erhebungen genutzt werden. Es ist folglich nur eine Annäherung an die optimale Lösung möglich.

53 Die durch die Auswahl des Verfahrens implizit gewählte Zielfunktion "Maximierung der relativen Entropie" induziert ein nicht-lineares Optimierungsproblem.

54 Unter "globalem Minimum" ist die minimale Abweichung zu den vorgegebenen Randinformationen zu verstehen.

### 3.2 Das Prinzip der "Maximalen relativen Entropie"<sup>55</sup>

Die Konstruktion synthetischer Mikrodaten kann als Schätzung einer Tabelle höherer Ordnung aus Tabellen niedrigerer Ordnung verstanden werden (vgl. Arminger/Lijphart/Müller 1981, S. 275). Geht man von der a priori-Verteilung  $q$  aus, dann ist nach dem "Prinzip der maximalen relativen Entropie" die beste Schätzung für die unbekannte, diskrete Verteilung  $p^0$  diejenige Verteilung  $p^* \in D_0$ , die die relative Entropie bezüglich der Verteilung  $q$  maximiert (vgl. Gyárfás/Quinke 1993, S. 151; Paaß 1988, S. 9ff; Snickars/Weibull 1977, S. 146).<sup>56</sup> Die Entropie einer Verteilung  $p$  ist definiert als:<sup>57</sup>

$$(8) \quad H(p) = - \sum_{i=1}^k p_i \ln p_i,$$

und die relative Entropie bezüglich einer beliebigen Verteilung  $q$  als:

$$(9) \quad H_q(p) = - \sum_{i=1}^k p_i \ln \frac{p_i}{q_i}.$$
<sup>58</sup>

Grundidee bei diesem Prinzip ist, daß von der Verteilung  $p^0$  Teilinformationen in Form von a priori-Informationen vorliegen und sonst keine weiteren Informationen bekannt sind ("man weiß etwas, aber nicht alles", vgl. Theil/Fiebig 1984, S. 5). Vor diesem Hintergrund ist es rational, die unbekannte Verteilung  $p^0$  durch eine angepaßte Verteilung  $\hat{p}$  zu schätzen, die so unbestimmt wie möglich bei Erfüllung der a priori-Informationen in Form von Nebenbedingungen ist.

Um diese Rationalität zu zeigen, sei angenommen, daß eine andere als die maximal unbestimmte Verteilung gewählt wird. Diese Verteilung enthält dann aber mehr Informationen als die a priori-Bedingungen vorgeben, was aber irrational

<sup>55</sup> Das Entropiemaß wurde von Shannon (1948) in die statistische Informationstheorie eingeführt. Eine ausführliche Diskussion dieses Prinzips findet sich in dem Sammelband von Levine/Tribus 1978.

<sup>56</sup> Gyárfás (1990, S. 56) bezeichnet das "Prinzip der maximalen Entropie" als theoretische Leitvorstellung für die Konstruktion synthetischer Mikrodaten überhaupt. Snickars/Weibull nennen das Prinzip "Minimale Information" und definieren den Term ohne das negative Vorzeichen. Einer Maximierung der relativen Entropie entspricht somit bei Snickars/Weibull eine Minimierung der Information.

<sup>57</sup> Vgl. Theil/Fiebig 1984, S. 5; vgl. auch Zellner/Highfield 1988, Shore/Johnson 1980, S. 26; Gray 1990, S.90.

<sup>58</sup> Für stetige Verteilungen ist die relative Entropie definiert als:

$$H_q(p) = - \int p(x) \ln \frac{p(x)}{q(x)} dx.$$

ist, da angenommen wurde, daß die Nebenbedingungen alle a priori-Informationen enthalten, die über die Verteilung  $p$  bekannt sind (vgl. Theil/Fiebig 1984, S. 5).

Die in (8) angegebene Beziehung mißt den Grad der Unbestimmtheit einer Verteilung  $p$ . Folglich sollte die Funktion ihr Maximum bei größtmöglicher Unbestimmtheit erreichen. Jaglom/Jaglom (1984, S. 340ff) zeigen, daß  $H(p)$  das Maximum bei  $p_i = \frac{1}{k}$  erreicht, wenn über die Verteilung  $p$  nur bekannt ist, daß es  $k$

mögliche Ausprägungen gibt und  $\sum_{i=1}^k p_i = 1$  (vgl. auch Theil/Fiebig 1984, S. 6).

Bei  $p_i = \frac{1}{k}$  handelt es sich gerade um die Gleichverteilung mit gleicher Wahrscheinlichkeit für jedes mögliche Ereignis; die Unbestimmtheit ist also maximal. Der Maximalwert ergibt sich dann zu:

$$(10) \quad H(p) = - \sum_{i=1}^k \frac{1}{k} \ln \frac{1}{k} = \ln k.$$

Aus der Beziehung (10) folgt, daß mit steigender Zahl  $k$  der möglichen Ausprägungen auch der Grad der Unbestimmtheit zunimmt.<sup>59</sup>

Die in (8) angegebene Funktion ist nicht willkürlich gewählt. Jaglom/Jaglom (1984, S. 104) beweisen, daß aus drei plausiblen Forderungen an die Funktion  $H$ , die in (8) angegebene Funktion bis auf einen konstanten Faktor eindeutig bestimmt ist. Es wird von der Funktion  $H$  gefordert, daß

- der Wert von  $H(p_1, \dots, p_k)$  sich durch Vertauschen der  $p_1, p_2, \dots, p_k$  nicht ändert,
- die Funktion  $H(p_1, \dots, p_k)$  stetig ist, d.h. ihr Wert ändert sich bei kleinen Veränderungen der  $p_1, p_2, \dots, p_k$  auch nur geringfügig, denn bei geringen Veränderungen der Wahrscheinlichkeiten darf sich auch der Grad der Unbestimmtheit nur wenig ändern
- und  $H$  der Beziehung genügt:

$$(11) \quad H(p_1, p_2, \dots, p_k) = H(p_1 + p_2, p_3, \dots, p_k) + (p_1 + p_2) H\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right)$$

---

<sup>59</sup> Für  $k = 1$  ist der Wert der Entropie Null, da das Ereignis  $p$  mit Sicherheit eintritt. Existiert ein  $p_i = 1$ ,  $i = \{1, \dots, k\}$ , dann sind alle  $p_j$  für  $i \neq j$  Null. Folglich muß  $H(p)$  ebenfalls Null werden, da mit Sicherheit das Ereignis  $p_i$  eintritt. In diesem Fall muß jedoch die Definition der Entropie um den Term  $0 \ln 0 := 0$  erweitert werden.

Letztere Forderung bedeutet, daß die in der Verteilung enthaltene Unsicherheit sinkt, wenn Wahrscheinlichkeiten aggregiert werden, und zwar um die Unsicherheit, die in der Verteilung der zusammengefaßten Ereignisse enthalten ist gewichtet mit der Wahrscheinlichkeit des zusammengefaßten Ereignisses (vgl. Gyárfás 1990, S. 57; Jaglom/Jaglom 1984, S. 105ff).<sup>60</sup>

Für das in Abschnitt III.3.1 formulierte Optimierungsproblem liegen aber zusätzliche Informationen über die Verteilung  $p$  in Form von linearen Randbedingungen und einer a priori-Verteilung  $q$  vor. In diesem Fall soll als beste Schätzung  $\hat{p}$  für die unbekannte Verteilung  $p^0$  diejenige Verteilung gelten, die bei Erfüllung der Nebenbedingungen die relative Entropie (vgl. Gleichung (9)) bezüglich  $q$  maximiert (vgl. Zellner/Highfield 1988, S. 197ff; Snickars/Weibull 1977, S. 145ff). Inhaltlich bedeutet dies: Der Grad der Unbestimmtheit der Verteilung  $\hat{p}$  soll nur um so viel geringer sein als der Grad der Unbestimmtheit der Verteilung  $q$ , wie die Randbedingungen dies erfordern. Ist die a priori-Verteilung eine Gleichverteilung, dann ist eine Maximierung der relativen Entropie gleichbedeutend mit einer Maximierung der Entropie (vgl. Abschnitt III.7).

Gyárfás (1990, S. 56f.) begründet seine Auffassung, daß im Prinzip der maximalen relativen Entropie der ideale Ansatz für die Konstruktion synthetischer Mikrodaten zu sehen ist, durch mehrere Argumente (vgl. auch Gyárfás/Quinke 1993, S. 152; Gyárfás/Quinke 1991, S. 3ff):

- Das Entropie-Prinzip stellt die einzige Methode dar, die sich eindeutig aus plausiblen "Konsistenz-Axiomen" herleiten läßt (vgl. Shore/Johnson 1980, S. 27ff).<sup>61</sup>
- Das Prinzip steht in enger Verbindung mit einem Grenzwertsatz: Die empirische Verteilung von unabhängigen diskreten Zufallszahlen mit der Verteilung  $q$  konvergiert bei Erfüllung der Restriktionen für wachsenden Stichprobenumfang gegen diejenige Verteilung, die unter den gegebenen Nebenbedingungen die relative Entropie bezüglich  $q$  maximiert (vgl. van Campenhou/cover 1981, S. 483).
- Viele bei der Analyse statistischer Zusammenhänge angenommene theoretische Verteilungen sind in Abhängigkeit von ihren Parametern entropiemaximale Verteilungen:

---

<sup>60</sup> Der Beweis, daß die resultierende Funktion eindeutig bestimmt ist, vereinfacht sich, wenn man zusätzlich fordert, daß die Funktion  $H\left(\frac{1}{k}, \dots, \frac{1}{k}\right)$  mit wachsendem  $k$  ebenfalls wächst. Dies ist gleichbedeutend mit einem zunehmenden Grad der Unbestimmtheit bei einer Zunahme der möglichen Ausprägungen.

<sup>61</sup> Es handelt sich hier nicht um die oben formulierten Forderungen, die die Entropie als Informationsmaß charakterisieren (vgl. Gyárfás/Quinke 1991, S. 3; Jaglom/Jaglom 1984, S. 105ff).

Wenn für eine stetige Verteilung nur positive Werte zugelassen werden und der Erwartungswert bekannt ist, dann ist die zugehörige entropiemaximale Verteilung die Exponentialverteilung.

Die Gleichverteilung ergibt sich als Schätzung einer Verteilung  $p_1, p_2, \dots, p_k$  über die keinerlei Information vorliegt ("Prinzip vom unzureichenden Grunde").

Für den Fall, daß von einer stetigen Verteilung nur Erwartungswert  $\mu$  und Varianz  $\sigma^2$  bekannt sind, ergibt sich als zugehörige entropiemaximale Verteilung die Normalverteilung  $N(\mu, \sigma^2)$ .

- Es besteht ein enger Zusammenhang zwischen Maximierung der Entropie und den log-linearen Modellen, so daß die statistischen Eigenschaften der entropiemaximalen Verteilung bekannt sind.
- Die Zielfunktion eignet sich für das hier interessierende Problem, da für festes  $q$  die (relative) Entropie eine stetige Funktion auf  $P$  (vgl. Gleichung (5)) ist und damit aufgrund der Kompaktheit von  $D_0$  ein maximierendes Element  $p^*$  existiert. Die Verteilung  $p^*$  ist eindeutig, denn die Funktion  $H_q(p)$  ist eine streng konkave Funktion auf  $P$ , und  $D_0$  ist eine konvexe Menge (vgl. Gyárfás/Quinke 1993, S. 151; ähnlich bei Blien/Graef 1991, S. 405).<sup>62</sup>

Verfahren der Entropiemaximierung (absolut oder relativ) haben bereits breite Anwendung in den Wirtschafts- und Sozialwissenschaften zur Schätzung von Verteilungen vor dem Hintergrund unvollständiger Information gefunden.<sup>63</sup> Die theoretischen Eigenschaften des Entropieprinzips einerseits und das sehr breite Anwendungsfeld andererseits lassen dieses Prinzip als geeignete Zielfunktion für die Konstruktion synthetischer Mikrodaten erscheinen. An die Diskussion des Prinzips der (relativen) maximalen Entropie schließt sich nun die Vorstellung und Analyse des zur Lösung des Optimierungsproblems aus Abschnitt III.3.1 eingesetzten Verfahrens an.

---

<sup>62</sup> Blien/Graef (1991) charakterisieren die Minimalstelle von Gleichung (5) durch die Karush-Kuhn-Tucker-Bedingungen (vgl. auch Collatz/Wetterling 1971, S. 111).

<sup>63</sup> Vgl. z.B.: Stäglin 1970; Graef/Blien 1989; Blien/Graef 1991; Merz 1983; Vassilion 1984; Jörnsten/Larsson/Lundgren/Migdalas 1990; Crouchly/Pickles/O'Farrell 1991; Nisan/Caveny 1988; Bellmann/Reinberg/Tessaring 1992, S. 13).

### 3.3 Simulated Annealing

#### 3.3.1 Darstellung des Verfahrens<sup>64</sup>

Zur Lösung des oben geschilderten Optimierungsproblems wird der Algorithmus "Simulated Annealing" verwendet. Andere gebräuchliche Bezeichnungen für diesen Algorithmus sind Monte-Carlo Annealing, Statistical Cooling, Probabilistic Hill Climbing, Stochastic Relaxation oder Probabilistic Exchange Algorithm (vgl. Laarhoven/Aarts 1987, S. 5).

Das Simulated Annealing wurde von Metropolis et. al. (1953, zit. n. Faigle/Schrader 1988) entdeckt und hat bis heute eine große Verbreitung erfahren. Anwendung hat das Verfahren in so unterschiedlichen Feldern wie der Simulation des Verhaltens idealer Gase bei Absenkung der Temperatur auf den absoluten Nullpunkt, der Berechnung guter Banken beim Fußballtoto oder der Konstruktion synthetischer Stichproben gefunden.<sup>65</sup>

Das Verfahren ist bereits erfolgreich zur Konstruktion synthetischer Mikrodaten eingesetzt worden. Auf Basis von Aggregatdaten der Steuerstatistik wurde eine Stichprobe von Steuerpflichtigen zur Simulation der Auswirkungen unterschiedlicher Steuergesetze erzeugt. Diese Datenbasis wird in verschiedenen Bundesministerien zur Schätzung der Kosten und Verteilungswirkungen von Gesetzesänderungen eingesetzt (vgl. Gyárfás 1990, Gyárfás 1988). Ein sehr breites Anwendungsspektrum und der bereits erfolgreiche Einsatz zur Konstruktion synthetischer Daten legen die Vermutung nahe, daß durch Einsatz einer Standardimplementierung des Simulated Annealing das sich hier stellende Problem lösen ließe. Faigle/Schrader (1988) zeigen jedoch in ihren Simulationsexperimenten, daß zufriedenstellende Ergebnisse - im Sinne von Auffinden der Optimallösung bzw.

---

<sup>64</sup> Vgl. zum folgenden: Laarhoven/Aarts 1987, S. 7ff; Aarts/Laarhoven 1985, S. 196ff; Aarts/Laarhoven 1989, S. 37ff; Aarts/Laarhoven 1987, S. 179ff; Gyárfás 1990, S. 42ff; Gyárfás/Quinke 1993, S. 152ff; Gyárfás/Quinke 1991, S. 4ff; Gyárfás 1988, S. 17ff; Paaß 1988, S. 7ff.

Die hier angegebene Literatur stellt nur einen kleinen Auszug aus einem großen Spektrum von Veröffentlichungen zur Theorie und Anwendung des Simulated Annealing dar. Eine einfache Darstellung findet sich bei Kuhn 1992.

<sup>65</sup> Die Eignung des Simulated Annealing zur Lösung kombinatorischer Optimierungsprobleme wurde erstmals von Kirkpatrick/Gelatt/Vecchi (1983) gezeigt. Die Idee, das Simulated Annealing zur Konstruktion synthetischer Stichproben zu nutzen, stammt von Paaß (1988).

Neben den zitierten Anwendungsfällen seien noch einige andere aufgeführt: Graphenkolourierung (vgl. Johnson/Aragon/McGeoch/Schevon 1991), Traveling Salesman Problem (vgl. Bonomi/Lutton 1984), Scheduling (vgl. Fleury 1991, Kuhn 1990, S. 173ff).

Annäherung an die Optimallösung<sup>66</sup> - nur bei einer problemadäquaten Implementation zu erwarten sind. "Zusammenfassend kann gesagt werden, daß das Simulated Annealing eine vielversprechende Methode zur computergestützten Bewältigung großer Optimierungsprobleme ist, dessen Erfolg bisher noch entscheidend vom Geschick des einzelnen Implementierers abhängt" (Faigle/Schrader 1988, S. 263). Die Konvergenz und die Konvergenzgeschwindigkeit hängen nach Auffassung der Autoren entscheidend von den gewählten Modellparametern ab. Die umfangreichen Simulationsexperimente lassen erwarten, daß die hier gewählte Implementation sowohl die Konvergenz gegen das Kostenminimum als auch die Konvergenz gegen die entropiemaximale Verteilung sicherstellt. Gegenwärtig ist ungeklärt, ob Simulated-Annealing-Algorithmen sich zu einem benutzerunabhängigen automatisierten Werkzeug entwickeln können (vgl. ebenda, S. 259).<sup>67</sup>

Simulated Annealing Algorithmen bestehen aus drei Teilen:

### 1. Generierungsmechanismus

Zu einem gegebenen Zustand  $x$  wird stochastisch ein potentieller Nachfolgezustand  $y \in N(x) \subseteq S$  erzeugt. Dabei bezeichnet  $S$  die Menge aller zulässigen Zustände und  $N(x)$  die Menge aller Nachbarn von  $x$ . In unserem Fall sind die Zustände  $x$  und  $y$  Stichproben von Unternehmen. Die Menge  $N(x)$  beinhaltet demnach alle Stichproben (als Nachbarn von  $x$ ), die sich höchstens in den Variablenwerten eines Falles unterscheiden.

### 2. Akzeptanzkriterium

Ein im Generierungsschritt erzeugter Zustand kann zu einer Verbesserung ( $\Delta C_{xy} := C(x) - C(y) \geq 0$ ) oder zu einer Verschlechterung ( $\Delta C_{xy} < 0$ ) führen. Die Funktion  $C(x)$  bzw.  $C(y)$  mißt den Abstand zu den vorgegebenen Randbedingungen: Aus der mittels des Generierungsmechanismus erzeugten Stichprobe werden durch geeignete Berechnungsvorschriften die zu dieser Stichprobe gehörenden Aggregatinformationen (ein- und mehrdimensionale Häufigkeitsverteilungen, Mittelwerte, Summen von Variablenwerten) berechnet und mit den vorgegebenen Randinformationen verglichen. Ist der Abstand des Nachfolgezustandes  $y$  zu den Randbedingungen geringer als der Abstand des Zustandes  $x$ , handelt es sich um eine Verbesserung. Verbesserungen werden immer akzeptiert, aber in einem bestimmten Umfang auch Verschlechterungen. Die Akzeptanz von Verschlechterungen soll verhindern, daß das Verfahren im Laufe des Simulationsprozesses in einem lokalen Minimum der Kostenfunktion hängenbleibt. Die Zahl der akzeptierten Ver-

---

<sup>66</sup> Die Simulationen werden für Probleme durchgeführt, bei denen sich die Optimallösung mittels anderer Verfahren bestimmen läßt.

<sup>67</sup> Die mittels Simulated Annealing erzeugten Ergebnisse sind bei einer entsprechenden problemspezifischen Implementation nicht subjektiv, sondern konvergieren sowohl gegen das Kostenminimum als auch gegen das Entropiemaximum.

schlechterungen wird durch eine Akzeptanzwahrscheinlichkeit gesteuert, die von  $\Delta C_{xy}$  und einem Kontrollparameter  $T > 0$  ("Temperatur") abhängt.<sup>68</sup>

### 3. Kontrollschema

Hier wird der "Kühlplan" festgelegt, d. h. die Werte des Kontrollparameters  $T$  für den Iterationsprozeß werden bestimmt.

Die Erzeugung der möglichen Nachfolger  $y \in N(x)$  zu einem gegebenen Zustand  $x$  erfolgt mittels der Übergangswahrscheinlichkeiten  $g_{xy}$ .<sup>69</sup> Für die  $g_{xy}$  wird gefordert:

$$(12) \quad g_{xy} > 0 \quad \forall y \in N(x); \quad \sum_{y \in N(x)} g_{xy} = 1 \quad 70.$$

Die Akzeptanzwahrscheinlichkeit sei gegeben durch:<sup>71</sup>

$$(13) \quad a_{xy}(\Delta C_{xy}, T) = \min \left\{ 1, e^{\frac{\Delta C}{T}} \right\}.$$

Für die Wahl dieser Funktion sprechen einige wünschenswerte Eigenschaften (vgl. z. B.: Gyárfás/Quinke 1991, S. 5):

- Verbesserungen werden immer akzeptiert, da für  $(\Delta C_{xy} \geq 0) \Rightarrow e^{\frac{\Delta C}{T}} \geq 1$  und daher  $a_{xy}(\Delta C_{xy}, T) = 1$ .
- Für gegebenes  $T > 0$  werden Verschlechterungen um so eher akzeptiert, je geringer sie sind (je größer  $\Delta C_{xy}$  ist, mit  $\Delta C_{xy} < 0$ ).
- Für  $\Delta C_{xy} < 0$  ist die Akzeptanzwahrscheinlichkeit eine monotone Funktion von  $T$  und es gilt:

<sup>68</sup> Die Bezeichnung des Kontrollparameters als "Temperatur" ist auf die ursprüngliche Verwendung des Verfahrens zur Simulation des Verhaltens idealer Gase bei Absenken der Temperatur auf den absoluten Nullpunkt zurückzuführen.

<sup>69</sup> Da zwei benachbarte Zustände sich nur in den Variablenwerten eines Falles unterscheiden und ein Fall in dieser Implementation des Algorithmus mittels gleichverteilter Zufallszahlen ausgewählt wird, sind die  $g_{xy}$  durch dieses Auswahlverfahren und die zur Veränderung der Variablenwerte genutzten Verteilungen bestimmt.

<sup>70</sup> Aus der Definition von  $N(x)$  folgt:  $x \in N(x)$ ;  $g_{xx} > 0$ .

<sup>71</sup> Diese Akzeptanzwahrscheinlichkeit ist die in der Literatur dominierende Akzeptanzwahrscheinlichkeit; vgl. z. B.: Anily/Federgruen 1987b, S. 661ff; de Soete 1988, S. 332f.; Sofianopoulou 1992, S. 329; Bohachevski/Johnson/Stein 1986, S. 211; Cerný 1985, S. 43; Anily/Federgruen 1987a, S. 873; Lundy/Mees 1986, S. 113.

$$(14) \quad \lim_{T \rightarrow \infty} a_{xy}(\Delta C_{xy}, T) = 1; \quad \lim_{T \rightarrow 0} a_{xy}(\Delta C_{xy}, T) = 0; \quad \Delta C_{xy} \text{ gegeben.}$$

Durch die Gültigkeit der Beziehung (14) kann die Zahl der akzeptierten Verschlechterungen im Laufe des Iterationsprozesses reduziert werden, indem der Kontrollparameter  $T$  abgesenkt wird. Die Entscheidung, ob ein Nachfolgezustand akzeptiert wird, ergibt sich aus dem Vergleich einer beliebigen gleichverteilten Zufallszahl aus dem Intervall  $[0,1]$  mit der berechneten Akzeptanzwahrscheinlichkeit. Ist die aus  $[0,1]$  gezogene Zufallszahl kleiner oder gleich  $a_{xy}(\Delta C_{xy}, T)$ , dann wird die Veränderung akzeptiert. Hieraus ergibt sich, daß Verbesserungen immer akzeptiert werden, da für  $\Delta C_{xy} > 0$  (Verbesserung) die Größe  $a_{xy}(\Delta C_{xy}, T) = 1$  ist.<sup>72</sup>

In einer Pseudo-Programmiersprache läßt sich der Algorithmus wie folgt formulieren:<sup>73</sup> Für eine gegebene Folge von Temperaturen  $(T_t)$ ,  $t = 1, 2, \dots$  (Kühlplan) sind folgende Anweisungen auszuführen:

Initialisierung:

- (1) Generiere mittels der a priori-Verteilung einen beliebigen Anfangszustand  $z \in S$ .
- (2) Setze Zähler = 1
- (3) Erzeuge zu dem Zustand  $z$  einen Nachfolgezustand  $x \in N(z)$ .

Iteration:

- (4) Berechne  $\Delta C_{zx} = C(z) - C(x)$ .
- (5) Berechne  $a_{zx}(\Delta C_{zx}, T) = \min \left\{ 1, e^{-\frac{\Delta C_{zx}}{T}} \right\}$ .
- (6) Generiere eine gleichverteilte Zufallszahl  $r$  mit  $0 \leq r \leq 1$ .
- (7) Wenn  $r > a_{zx}(\Delta C_{zx}, T)$ , gehe zu (9).
- (8) Akzeptiere  $x$  als neuen aktuellen Zustand;  $z := x$ .
- (9) Setze Zähler = Zähler + 1.

---

<sup>72</sup> Zum Problem der Erzeugung gleichverteilter Zufallszahlen, bzw. Zufallszahlen beliebiger Verteilung vgl. z.B.: Hengartner/Theodorescu 1978, S. 25ff, Witte 1973, S. 143ff, Krüger 1975, S. 112ff. Die bei der Programmierung des Algorithmus verwendete Version des Turbo Pascal Compilers enthält eine Prozedur zur Erzeugung gleichverteilter Pseudozufallszahlen.

<sup>73</sup> Vgl. z.B. Faigle/Schrader 1988, S. 261, Paaß 1988, S. 7f.

(10) Gehe zu (3).

Die Implementierung des Verfahrens zur Lösung eines spezifischen Optimierungsproblems setzt einige Spezifikationen voraus. Eglese (1990, S. 274f) unterscheidet zwischen generellen und problemspezifischen Spezifikationen:

### Generelle Spezifikationen

Die folgenden Auswahlentscheidungen müssen bei jeder Implementation des Algorithmus getroffen werden:

- Bestimmung des Startwerts für den Kontrollparameter  $T$
- Es muß eine Funktion gewählt werden, die den Kontrollparameter im Laufe des Iterationsprozesses steuert.
- Die Zahl der Iterationsschritte für jede Temperatur  $T_1, T_2, \dots$  muß festgelegt werden.
- Ein Stoppkriterium, bei dem der Algorithmus abbrechen soll, ist zu definieren.

### Problemspezifische Spezifikationen

- Das Optimierungsproblem muß "klar" formuliert sein, damit eine Menge zulässiger Lösungen existiert (vgl. Abschnitt III.3.1).
- Von zentraler Bedeutung ist die problemadäquate Spezifikation der Nachbarschaftsstruktur zu einem gegebenen Zustand  $x$  (vgl. Cerný 1985).<sup>74</sup>
- In Abhängigkeit von der Problemstellung muß die Funktion  $C_{xy}$  gewählt werden.<sup>75</sup>

Die hier gewählte Implementation und die zugehörigen Spezifikationen werden ausführlich im Zusammenhang mit der Implementationsbeschreibung dargestellt. Zunächst erfolgt eine theoretische Analyse des Verfahrens, aus der sich einige Forderungen für die oben skizzierten Spezifikationen ergeben, wenn einerseits die Konvergenz zum globalen Optimum der Kostenfunktion und andererseits die Konvergenz gegen die entropiemaximale Verteilung theoretisch sichergestellt sein soll.

---

<sup>74</sup> Vgl. dazu ausführlich Abschnitt IV.4.2.2.

<sup>75</sup> In der Anwendung zur Erzeugung von synthetischen Stichproben mißt  $C_{xy}$  die Abweichung von den vorgegebenen Randbedingungen.

### 3.3.2 Konvergenz des Simulated Annealing zum globalen Kostenminimum

Der Beweis zur Konvergenz des Simulated Annealing basiert auf der Theorie homogener und inhomogener Markovketten. Die Beweise der folgenden Sätze finden sich in der angegebenen Literatur.

Eine durch das Simulated Annealing erzeugte Folge von Zuständen läßt sich als Markovprozess darstellen. Die Übergangswahrscheinlichkeiten von einem beliebigen Zustand  $x \in S$  zu einem Zustand  $y \in S$  sind gegeben durch:

$$(15) \quad p_{xy}(T_t) = \begin{cases} 0 & \text{für } y \notin N(x) \\ g_{xy} a_{xy}(T_t) & \text{für } y \in N(x); y \neq x \\ 1 - \sum_{\substack{z \in N(x), \\ z \neq x}} p_{xz}(T_t) & ; x = y \end{cases}$$

$g_{xy}$ <sup>76</sup> steht hier für den Generierungsmechanismus  $G$ , der zunächst genauer spezifiziert werden muß. Für festes  $T$  erzeugt der Algorithmus eine homogene Markovkette.<sup>77</sup> Die folgenden Ausführungen zeigen den Zusammenhang zwischen dem Parameter  $T$  und den zu minimierenden Kosten. Für die Konstruktion der synthetischen Unternehmensstichprobe wird folgender Generierungsmechanismus verwendet:

- a) Zunächst wird eine Anfangsstichprobe ( $t = 0$ ) erzeugt, d.h. jedem Fall (Unternehmen)  $i = 1, \dots, n$  wird durch eine a priori-Verteilung  $q = (q_1, \dots, q_k)'$  ein Wert  $x_i(0) \in \{1, \dots, k\}$  zugewiesen.
- b) In jedem Iterationsschritt  $t \geq 1$  wird mit der Wahrscheinlichkeit  $1/n$  ein Fall ausgewählt und einer Variablen dieses Falls wird mittels der Verteilung  $q$  ein Wert  $x_i(t)$  zugewiesen.

---

<sup>76</sup> Bei  $g_{xy}$  handelt es sich auch um eine Wahrscheinlichkeit, die jedoch unabhängig von der Akzeptanzwahrscheinlichkeit ist.

<sup>77</sup> Eine Markovkette heißt homogen, wenn die Übergangswahrscheinlichkeiten zeitunabhängig sind (vgl. Heller/Lindenberg/Nuske/Schriever 1978, S. 23). Da  $T$  für eine bestimmte Zahl von Iterationen als konstant angenommen wird, ist diese Bedingung im vorliegenden Fall erfüllt.

Die Übergangswahrscheinlichkeiten zwischen zwei Zuständen  $x$  und  $y$  ergeben sich damit zu:<sup>78</sup>

$$(16) \quad g_{xy} = \begin{cases} 0 & \text{für } y \notin N(x) \\ \frac{1}{n} q_{y_i} & \text{für } y \in N(x); y \neq x \\ \frac{1}{n} \sum_{i=1}^n q_{x_i}, & x = y, \end{cases}$$

wobei  $N(x)$  die Menge aller Zustände bezeichnet, die sich höchstens in den Variablenwerten eines Falles unterscheiden.

Bezeichne  $Q_x(t)$  die Wahrscheinlichkeit, daß nach dem Iterationsschritt  $t$  ein Zustand  $x = (x_1, \dots, x_n)'$  vorliegt. Die zu  $G$  (aus Gleichung 16) gehörende Markovkette hat dann folgende Eigenschaften (vgl. Gyárfás/Quinke 1993, S. 155):<sup>79</sup>

Satz 1: Die durch  $G$  erzeugte Folge von Zuständen  $\{x(t)\}$  bildet eine homogene, aperiodische und irreduzible Markovkette, welche folgende Bedingungen erfüllt:

$$(17) \quad Q_x(t) = Q_x(0) =: Q_x \quad \forall x \in S, t = 1, 2, \dots$$

$$(18) \quad Q_x g_{xy} = Q_y g_{yx} \quad \forall x, y \in S. \text{ }^{80}$$

Eine Markovkette wird als homogen bezeichnet, wenn die Übergangswahrscheinlichkeiten unabhängig vom Iterationsschritt sind. Diese Eigenschaft besitzen die in Gleichung (16) definierten Übergangswahrscheinlichkeiten. Aperiodisch heißt eine Markovkette, wenn die Wahrscheinlichkeit, den gleichen Zustand nach einem Übergang erneut zu erhalten, größer als Null ist ( $g_{xx} > 0$ ). Von einer irreduziblen Markovkette wird gesprochen, falls jeder Zustand  $y \in N(x)$  in einer endlichen Zahl von Schritten von einem Zustand  $x \in N(x)$  erreichbar ist. Die Kette aus Gleichung (16) besitzt diese Eigenschaft, da sich zwei Zustände

<sup>78</sup>  $g_{xy}$  bezeichnet den in (15) nicht genauer spezifizierten Generierungsmechanismus. Um die  $p_{xy}(T_t)$  zu erhalten, muß  $g_{xy}$  noch mit  $a_{xy}(T_t)$  multipliziert werden, da eine Veränderung nur mit der Wahrscheinlichkeit  $a_{xy}(T_t)$  akzeptiert wird.

<sup>79</sup> Zum Beweis des Satzes vgl. Gyárfás/Quinke 1991, S. 7f.

<sup>80</sup> Diese Eigenschaft wird zum Beweis von Satz 2 benötigt.

$x, y \in N(x)$  in den Variablenwerten von  $n'$  Fällen unterscheiden ( $0 \leq n' \leq n$ ). Folglich ist  $y$  von  $x$  aus in einer endlichen Zahl von Schritten erreichbar.

In Satz 1 werden zunächst die Eigenschaften der Übergangswahrscheinlichkeiten  $g_{xy}$  aus Gleichung (16) untersucht, die im Unterschied zu den  $p_{xy}(T_t)$  aus Gleichung (15) unabhängig von der Akzeptanzwahrscheinlichkeit sind. Besitzt auch die durch die  $p_{xy}(T_t)$  gegebene Markovkette (für festes  $T$ ) die Eigenschaften aus Satz 1, dann kann durch Anwendung zentraler Sätze aus der Theorie der Markovketten gezeigt werden, daß die zu Gleichung (15) gehörende Markovkette eine stationäre Verteilung hat.

Es soll nun wieder die in (15) definierte Markovkette betrachtet werden.<sup>81</sup> Da die Akzeptanzwahrscheinlichkeit immer größer als Null ist, folgt aus Satz 1, daß für festes  $T$  auch die Kette aus (15) homogen, aperiodisch und irreduzibel ist. Es sei  $p_x^t(T)$  die Wahrscheinlichkeit von Zustand  $x$  nach dem Iterationsschritt  $t$ . Dann gilt (vgl. Gyárfás/Quinke 1991, S. 8):

Satz 2: Für eine durch das Simulated Annealing erzeugte Markovkette gilt:<sup>82</sup>

$$(19) \quad \forall x \in S \quad \lim_{t \rightarrow \infty} p_x^t(T) = \pi_x(T)$$

$$(20) \quad \pi_x(T) = K(t) Q_x e^{\left(\frac{-C(x)}{T}\right)} \quad \text{mit } K(T) = \frac{1}{\sum_{y \in S} Q_y e^{\left(\frac{-C(y)}{T}\right)}}$$

Für die zu Gleichung (15) gehörende Markovkette zeigt Satz 2, daß für eine große Zahl von Iterationsschritten ( $t \rightarrow \infty$ ) und konstantem Kontrollparameter  $T$  die  $p_x^t(T)$  gegen die stationären Wahrscheinlichkeiten  $\pi_x(T)$  mit der in Satz 2 angegebenen Gestalt konvergieren, die resultierende Verteilung also eindeutig bestimmt ist. Diese Aussage impliziert zunächst nur, daß die durch das Simulated Annealing erzeugte Markovkette eine stationäre Verteilung für konstantes  $T$  besitzt. Zu jedem  $T$  gehört demnach für  $t \rightarrow \infty$  eine stationäre Verteilung, zu der ein Kostenfunktionswert ermittelt werden kann.

81 Diese Markovkette entspricht der in dieser Arbeit gewählten Implementation des Simulated Annealing.

82 Diese Wahl von  $K(t)$  garantiert  $\sum_{x \in S} \pi_x(T) = 1$ . Vgl. Gyárfás 1990, S.47, Gyárfás/Quinke 1993, S. 156; Mitra/Romeo/Sangiiovanni-Vincentelli 1986, S. 753. Zum Beweis vgl. Gyárfás/Quinke 1991, S. 8f. Die Autoren benötigen für den Beweis jedoch nicht die Annahme  $g_{xy} = g_{yx}$ , d.h. es werden geringere Erfordernisse an die Nachbarschaftsstruktur gestellt.

Die bisherigen Ausführungen zeigen, daß für ein gegebenes und konstantes  $T$  sowie für eine hinreichend lange Laufzeit ( $t$  sehr groß) die Wahrscheinlichkeit eines bestimmten Zustands  $y$  unabhängig von der Anfangsverteilung wird. Der so erreichte Zustand approximiert  $\pi_x(T)$ . Wenn sich der Prozeß bereits im Gleichgewicht befindet, dann verändert sich die Wahrscheinlichkeit nicht, d.h. der Prozeß ist stationär (vgl. Gyárfás 1990, S.48).

Das Ergebnis von Gyárfás/Quinke (1993) und Gyárfás (1990) ist völlig identisch mit dem von Mitra/Romeo/Sangiovanni-Vincentelli (1986, S. 752ff.) außer, daß dort die Bezeichnungen anders gewählt sind und die stationären Wahrscheinlichkeiten nicht explizit angegeben sind, da der Generierungsmechanismus nicht vorgegeben wird (vgl. auch Laarhoven 1988, S. 14ff). Mitra/Romeo/Sangiovanni-Vincentelli benötigen jedoch für den Beweis die Bedingung  $g_{xy} = g_{yx}$ , das heißt, die Wahrscheinlichkeit, einen Zustand  $y$  als Nachfolgezustand zu  $x$  zu generieren, muß gleich der Wahrscheinlichkeit sein, von Zustand  $y$  nach Zustand  $x$  zu gelangen. In der hier gewählten Implementation ist diese Symmetriebedingung jedoch nicht erfüllt. Hat ein Unternehmen nur 20 Beschäftigte, dann ist die Wahrscheinlichkeit, diesem Unternehmen 1.000 und mehr Beschäftigte zuzuweisen, sehr gering, da bei der Generierung der Nachfolgezustände die empirische Verteilung der Unternehmen nach Betriebsgrößenklassen zugrundegelegt wird. Umgekehrt ist die Wahrscheinlichkeit, diesem Unternehmen bei gegebener Beschäftigtenzahl von 1.000 und mehr nur 20 Beschäftigte zuzuweisen, wesentlich größer. Der Beweis der Konvergenz ohne die Symmetrieannahme von Gyárfás/Quinke (1991) ist daher für die hier gewählte Implementation von zentraler Bedeutung.

Bisher wurde nur beschrieben, unter welchen Bedingungen bei festem  $T$  eine stationäre Verteilung existiert. Es muß noch gezeigt werden, daß die Wahrscheinlichkeit, einen global optimalen Zustand zu erreichen, bei Absenkung der Temperatur gegen eins konvergiert (vgl. Gyárfás/Quinke 1993, S. 156).<sup>83</sup>

Satz 3: Sei  $S$  die Menge der Zustände und  $S^* = \{x \in S \mid C(x) \leq C(y) \forall y \in S\}$  die Menge aller global optimalen Zustände. Dann gilt:

$$\lim_{T \rightarrow 0} p_T(x \in S^*) = 1.$$

Die Menge  $S^*$  beschreibt die Menge aller kostenoptimalen Zustände. Satz 3 postuliert, daß die Wahrscheinlichkeit, durch das Simulated Annealing - bei einer Absenkung des Kontrollparameters gegen Null - kostenminimale Zustände zu

---

<sup>83</sup> Zum Beweis des Satzes vgl. Gyárfás/Quinke 1991, S.9.  $p_T$  (in Satz 3) bezeichnet die Wahrscheinlichkeit in Abhängigkeit vom Kontrollparameter  $T$ , daß der Zustand  $x$  in der Menge aller kostenoptimalen Zustände enthalten ist.

erzeugen, gleich eins ist. Zunächst wurde gezeigt, daß für festes  $T$  durch das Simulated Annealing eine Markovkette erzeugt wird, welche die Eigenschaften aus Satz 2 besitzt. Zum Beweis des Satzes greifen Gyárfás/Quinke (1991, S. 9) auf die stationären Wahrscheinlichkeiten aus Satz 2 zurück und zeigen, daß für alle  $x \notin S$   $\lim_{T \rightarrow 0} \pi_x(T) = 0$  gilt. Daraus ergibt sich, daß sich die durch das Simulated Annealing ergebenden stationären Verteilungen kostenoptimale Verteilungen sind. Der Beweis von Gyárfás/Quinke setzt voraus, daß vor jeder Absenkung des Kontrollparameters  $T$  die stationäre Verteilung  $\pi_x(T)$  erreicht wird, da die grundlegende Idee des Beweises das Verhalten der  $\pi_x(T)$  bei  $T \rightarrow 0$  ist. In dieser Anwendung des Simulated Annealing wird der Kontrollparameter über eine große Zahl von Iterationen konstant gehalten, um so eine Annäherung an die stationäre Verteilung sicherzustellen.

Die hier vorgestellten Sätze zur Konvergenz des Simulated Annealing gegen einen global optimalen Zustand basieren auf der Modellierung durch homogene Markovketten und sind damit komparativ-statischer Natur. Wie oben beschrieben, wird in der hier implementierten Anwendung der Temperaturparameter verändert, das heißt, die zu dem gesamten Iterationsprozeß gehörige Markovkette ist inhomogen. Die Übertragung der Analyse homogener Markovketten auf den inhomogenen Fall setzt voraus, daß der Temperaturparameter solange konstant gehalten wird, bis die zugehörige stationäre Verteilung erreicht ist. Folglich muß sich eine Analyse anschließen, für welche "Kühlpläne" die Konvergenz zum globalen Optimum erfolgt.

### 3.3.3 Kühlpläne

Nach Laarhoven/Aarts (1987, S. 57) werden gewöhnlich Simulated-Annealing-Algorithmen verwendet, in denen eine Folge homogener Markovketten endlicher Länge für abnehmende Werte des Kontrollparameters erzeugt werden. Dazu sind folgende Parameter zu spezifizieren (vgl. auch 3.3.1):

1. der Anfangswert des Kontrollparameters  $T_0$ ,
2. der Endwert des Kontrollparameters  $T_e$
3. die Länge der Markovkette zu jedem Kontrollparameter  $T_k$
4. eine Regel, wie der aktuelle Wert des Kontrollparameters  $T_k$  verändert werden soll, also eine Antwort auf die Frage: Welche Gestalt hat bei gegebenen  $T_k$  der Wert  $T_{k+1}$  mit den Minimalanforderungen  $T_{k+1} \leq T_k$ ;  $k = 1, 2, \dots$  und  $\lim_{k \rightarrow \infty} T_k = 0$ .

Die so festgelegten Parameter bilden den Kühlplan für den Iterationsprozeß.

Wie oben bereits erwähnt, setzt die Modellierung des inhomogenen Falles mit Hilfe homogener Markovketten voraus, daß die Zahl der Schritte zwischen suk-

zessiven Änderungen von  $T$  so groß ist, daß jeweils ein Gleichgewicht (die stationäre Verteilung) erreicht wird. Im inhomogenen Fall spielt die Folge  $\{T_k\}$  der Kontrollparameter eine zentrale Rolle. Verschiedene Autoren haben Bedingungen für  $\{T_k\}$  formuliert, die eine Konvergenz zum globalen Optimum garantieren (vgl. Mitra/Romeo/Sangiovanni-Vincentelli 1986, S. 749ff; Hajek 1985, S. 313f. und S. 316f.; Aarts/Laarhoven 1987, S. 185; Aarts/Laarhoven 1989, S. 40).

Die theoretisch optimalen Kühlpläne sind für praktische Zwecke meist unbrauchbar, weil die Temperatur viel zu langsam abgesenkt wird, so daß der Iterationsprozeß sehr viel Zeit in Anspruch nimmt. Eine zu schnelle Absenkung der Temperatur kann die Konvergenz zum globalen Kostenminimum verhindern. In der Praxis werden daher Kühlpläne verwendet, die auf heuristischen Überlegungen basieren und deren Güte experimentell beurteilt wird.

Laarhoven/Aarts (1987, S. 55ff) diskutieren verschiedene Ansätze zur Bestimmung eines Kühlplans. Für die Konstruktion vieler Kühlpläne ist das Konzept des "Quasi-Gleichgewichts" zentral.

Sei  $L_k$  die Länge der  $k$ -ten Markovkette. Der Simulated-Annealing-Algorithmus befindet sich für  $T_k$  im Quasi-Gleichgewicht, falls " $a(L_k, T_k)$  nahe bei  $\pi_x(T_k)$ " ist, wobei  $a(L, T)$  die Wahrscheinlichkeitsverteilung nach  $L$  Schritten und  $\pi_x(T_k)$  die stationäre Verteilung der homogenen Markovkette für den Kontrollparameter  $T_k$  ist. Die exakte Definition des Begriffs "nahe bei" unterscheidet im wesentlichen verschiedene Kühlpläne voneinander.

Grundsätzlich basiert die Konstruktion eines Kühlplans auf folgenden Überlegungen:

- Für  $T_k \rightarrow \infty$  ist die stationäre Verteilung gegeben durch die a priori- Verteilung  $q$  auf der Menge der Konfigurationen  $S$  (vgl. Gyárfás/Quinke 1993, S. 160). Anfänglich kann ein Quasi-Gleichgewicht also dadurch erreicht werden, daß der Anfangswert von  $T$ ,  $T_0$  so gewählt wird, daß fast alle Übergänge akzeptiert werden. In diesem Fall treten alle Konfigurationen mit der Wahrscheinlichkeit  $q_i$  auf, d. h. es resultiert die Verteilung  $q$  und damit  $\pi_x(T_0)$ .
- Die Wahl eines Stopkriteriums basiert auf der Überlegung, daß die Ausführung des Algorithmus beendet werden kann, wenn die erwartete Verbesserung von  $\Delta C_{xy}$  bei einer weiteren Ausführung des Algorithmus klein ist.
- Die Länge  $L_k$  der  $k$ -ten Markovkette und die Transformationsregel zur Veränderung von  $T_k$  in  $T_{k+1}$  sind durch das Konzept des Quasi-Gleichgewichts sehr eng miteinander verbunden. Es ist evident, daß eine große Absenkung von  $T_k$  auf  $T_{k+1}$  im Vergleich zu einer langsamen Absenkung eine größere Zahl von Iterationsschritten notwendig macht, um das Gleichgewicht bei  $T_{k+1}$  wieder herzustellen. Es ergibt sich demnach ein trade-off zwischen ei-

ner schnellen Abnahme von  $T_k$  und kleinen Werten von  $L_k$ . In den meisten Fällen entscheidet man sich für kleine Abnahmen von  $T_k$ , um extrem lange Ketten zu vermeiden.

Laarhoven/Aarts (1987, S. 59ff) fassen die in der Literatur beschriebenen Kühlpläne zusammen. Die wichtigsten Ergebnisse werden im folgenden dargestellt.

### Anfangswert des Kontrollparameters

Der Anfangswert von  $T$  wird so festgelegt, daß nahezu alle Übergänge akzeptiert

werden, daß heißt  $T_0$  ist so groß, daß  $e^{\left(\frac{C(x)-C(y)}{T_0}\right)} \approx 1$  für fast alle  $x$  und  $y$ <sup>84</sup>.

Kirkpatrick/Gelatt/Vecchi (1983) schlagen folgende Regel vor: Wähle einen großen Wert für  $T_0$  und führe einige Übergänge durch. Falls das Akzeptanzverhältnis  $\chi$ , definiert als die Zahl der akzeptierten Übergänge dividiert durch die Zahl der vorgeschlagenen Übergänge, kleiner als ein vorgegebener Wert  $\chi_0$  (bei Kirkpatrick et. al.  $\chi_0 = 0,8$ ) ist, dann verdopple den aktuellen Wert von  $T_0$ . Führe diese Prozedur so lange durch, bis das beobachtete Akzeptanzverhältnis den Wert  $\chi_0$  übersteigt.

Diese Regel wurde von einigen Autoren verfeinert. Johnson et. al. (1991) bestimmen  $T_0$ , indem sie zunächst die durchschnittliche Kostenveränderung  $\overline{\Delta C}^{(t)}$  für eine bestimmte Anzahl von zufälligen Übergängen bestimmen und dann  $T_0$  aus der folgenden Beziehung berechnen:

$$(21) \quad \chi_0 = e^{\left(\frac{-\overline{\Delta C}^{(t)}}{T_0}\right)} \Leftrightarrow T_0 = -\frac{\Delta C^{(t)}}{\ln \chi_0} = \frac{\overline{\Delta C}^{(t)}}{\ln\left(\frac{1}{\chi_0}\right)}$$

Nahezu gleiche Ansätze zur Bestimmung von  $T_0$  verwenden Lundy/Mees (1986) und Aarts/Laarhoven (1989, S. 42).

---

<sup>84</sup> Man beachte hier, daß die Akzeptanzwahrscheinlichkeit  $a_{xy}$  vom Kontrollparameter  $T$  und von  $\Delta C_{xy}$  abhängt. Je kleiner  $\Delta C_{xy}$  (negativ) desto kleiner wird  $a_{xy}$ .

## Endwert des Kontrollparameters<sup>85</sup>

Ein Stopkriterium, welches den Endwert des Kontrollparameters determiniert, kann entweder durch Festlegen einer bestimmten Anzahl von Werten  $T_k$ , für die der Algorithmus ausgeführt wird, oder durch einen Abbruch des Algorithmus, falls die letzte Konfiguration aufeinanderfolgender Markovketten für eine vorbestimmte Zahl von Ketten identisch ist, bestimmt werden. Letztgenanntes Kriterium wurde von Johnson et. al. (1989) verfeinert: Es wird gefordert, daß das Akzeptanzverhältnis kleiner als ein gegebener Wert  $\chi_e$  ist. Lundy/Mees (1986, S. 122) bestimmen eine Schranke für den Endwert des Kontrollparameters  $T_e$ :

$$T_e \leq \frac{\varepsilon}{\ln(|S| - 1) - \ln \vartheta}$$

wobei  $\varepsilon$  den akzeptierten Fehler in der Lösung und  $\vartheta$  die Irrtumswahrscheinlichkeit beschreibt.<sup>86</sup>

Sowohl Aarts/Laarhoven (1985) als auch Otten/van Genniken (1984, zit. nach Laarhoven/Aarts 1987, S. 68) schlagen einen Endwert des Kontrollparameters vor, der auf dem Wachstum des durchschnittlichen Wertes der Kostenfunktion über eine Anzahl von Markovketten, die während der Ausführung des Algorithmus auftreten, basiert. Dazu bezeichne  $\bar{C}(T_k)$  diesen durchschnittlichen Kostenwert für die  $k$ -te Markovkette. Die Ausführung des Algorithmus ist beendet, wenn  $\bar{C}(T_k) - C_{\text{opt}}$  klein ist. Da man  $C_{\text{opt}}$  im allgemeinen nicht kennt, muß approximiert werden. Bei Laarhoven/Aarts (1987, S. 65) geschieht dies wie folgt: Falls  $T_k \ll 1$  erhält man

$$(22) \quad \bar{C}(T_k) - C_{\text{opt}} \approx T_k \left( \frac{\partial \bar{C}(T)}{\partial T} \right)_{T=T_k},$$

woraus sich  $T_e$  wie folgt bestimmen läßt:

$$(23) \quad \left( \frac{\partial \bar{C}(T)}{\partial T} \right)_{T=T_e} \frac{T_e}{\bar{C}(T_o)} < \varepsilon, \quad \text{für } \varepsilon > 0 \text{ klein.}$$

## Länge der Markovkette

Die einfachste Wahl für  $L_k$  (= Länge der  $k$ -ten Markovkette) ist ein Wert, der (polynomial) von der Größe des Problems abhängt;  $T_k$  ist damit unabhängig von

<sup>85</sup> Vgl. Laarhoven/Aarts 1987, S. 60 u. S. 64ff.

<sup>86</sup> Für eine genauere und etwas formaler Darstellung siehe Laarhoven/Aarts 1987, S. 64.

k. Andere Vorschläge für die Länge der Markovkette basieren auf dem intuitiven Argument, daß für jeden Wert  $T_k$  des Kontrollparameters eine Mindestzahl von Übergängen akzeptiert werden sollte, d.h.  $L_k$  ist so festgelegt, daß die Zahl der akzeptierten Übergänge wenigstens  $\eta_{\min}$  ( $\eta_{\min}$  ist eine beliebige, aber fest gewählte Zahl) beträgt.<sup>87</sup> Je näher  $T_k$  bei Null liegt, desto niedriger ist die Wahrscheinlichkeit, daß Übergänge akzeptiert werden und möglicherweise gilt:  $L_k \rightarrow \infty$  für  $T_k \rightarrow 0$ . Deshalb beschränkt man  $L_k$  mit Hilfe einer Konstanten  $L$  (gewöhnlich polynomial zur Problemgröße gewählt), um extrem lange Markovketten bei kleinen Werten von  $T_k$  zu vermeiden (vgl. Laarhoven/Aarts 1987, S. 61).<sup>88</sup>

### Abnahmeregel für den Kontrollparameter<sup>89</sup>

Eine häufig verwendete Abnahmeregel ist gegeben durch:

$$(24) \quad T_{k+1} = \alpha T_k, \quad k = 0, 1, 2, \dots$$

wobei  $\alpha$  eine Konstante ist, die kleiner aber nahe bei eins liegt ( $\alpha \in [0,5; 0,99]$ ). Eine andere Möglichkeit besteht darin, das Intervall  $[0; C_0]$  in eine feste Anzahl, etwa  $K$  Unterintervalle, aufzuteilen und  $T_k$ ,  $k = 1, \dots, K$  wie folgt zu wählen:

$$(25) \quad T_k = \frac{K-k}{K} T_0; \quad k = 1, \dots, K$$

(beachte: im ersten Fall ist  $\frac{T_{k+1}}{T_k} = \text{const}$  und im zweiten Fall ist  $T_k - T_{k+1} = \text{const}$ ).

Wie bereits oben erwähnt, wird der Kontrollparameter nur langsam verkleinert, damit die Länge der Markovketten bis zum Erreichen des Gleichgewichts nicht zu groß wird.<sup>90</sup>

<sup>87</sup> Akzeptierte Übergänge und generierte potentielle Nachfolgezustände können weit auseinanderliegen.

<sup>88</sup> Einen etwas ausgefeilteren Vorschlag zur Wahl der Länge der Markovketten findet sich bei Mitra/Romeo/Sangiovanni-Vingentelli 1986.

<sup>89</sup> Vgl. Laarhoven/Aarts 1987, S. 61f., S. 67ff).

<sup>90</sup> Laarhoven/Aarts (1987), Lundy/Mees (1986) und Otten/van Ginneken (1984) argumentieren, daß die stationären Verteilungen zweier aufeinander folgender Kontrollparameterwerte nahe zusammen liegen sollten:

$$\forall x \in S: \frac{1}{1+\delta} < \frac{\pi_x(T_k)}{\pi_x(T_{k+1})} < 1+\delta, \quad k = 0, 1, \dots; \delta \in \mathbb{R} \text{ klein.}$$

### 3.3.4 Kostenfunktion<sup>91</sup>

Die mittels des Simulated Annealing erzeugten Stichproben werden durch eine Kostenfunktion bewertet. Die durch die Randbedingungen vorgegebenen Verteilungen werden mit den zu der synthetischen Stichprobe gehörenden korrespondierenden Verteilungen verglichen, und die Abweichung von allen zugrundegelegten Randinformationen wird in einem Wert ausgedrückt. Je besser die Anpassung an die zu approximierenden Randverteilungen, desto niedriger soll der Wert der Kostenfunktion sein. Bei einer vollständigen Anpassung nimmt die Kostenfunktion den Wert Null an. Da die Wahl der Kostenfunktion von subjektiven Bewertungskriterien abhängt, ist eine Vielzahl von Spezifikationen möglich.

Paaß (1988, S. 2ff) schlägt Kostenfunktionen vor, die auf einer Analogie zur Maximum-Likelihood-Schätzung gründen. Nach der Maximum-Likelihood-Methode wird als Schätzer für den unbekannt Parameter derjenige Wert herangezogen, für den die Likelihoodfunktion ihren maximalen Wert annimmt. Geht man davon aus, daß eine vorgegebene Randverteilung auf einer Stichprobe basiert, dann können die Vorgaben in der Likelihoodfunktion als Stichproben behandelt werden. Als Parameter werden die auf Basis der synthetischen Stichprobe berechneten Werte eingesetzt.

Die Randinformation sei gegeben durch eine Tabelle mit  $k$  Zellen, die  $n_1, \dots, n_k$  Fälle enthalten. Es sei  $x$  eine synthetische Stichprobe mit  $n_1(x), \dots, n_k(x)$  Fällen in der durch die Randinformationen vorgegebenen Tabelle. Unterstellt man ein statisches Modell, bei dem  $k$  Ausprägungen mit den Wahrscheinlichkeiten  $p_1, \dots, p_k$  auftreten, dann ergibt sich für eine Stichprobe von  $n$  unabhängigen Beobachtungen die Wahrscheinlichkeit, daß  $n_1$  Fälle die erste Ausprägung,  $n_2$  Fälle die zweite Ausprägung etc. aufweisen aus der Multinomialverteilung. Es gilt:

$$(26) \quad p(n_1, \dots, n_k | p_1, \dots, p_k) = \frac{n!}{n_1! n_2! \dots n_k!} p_1^{n_1} \dots p_k^{n_k}.$$

Gesucht sind folglich die  $p_1, \dots, p_k$ , die die logarithmische Likelihoodfunktion maximieren. Die  $p_1, \dots, p_k$  werden so berechnet, daß die Wahrscheinlichkeit der vorgegebenen Randverteilung am größten ist:

$$(27) \quad L = \log(n!) - \sum_{i=1}^k \log(n_i!) + \sum_{i=1}^k n_i \log(p_i).$$

---

Die Autoren verwenden diese Gleichung als Ausgangspunkt, entwickeln aber unterschiedliche Abnahmeregeln (vgl. Laarhoven/Aarts 1987, S. 68).

91 Vgl. hierzu insbesondere Gyárfás 1990, S. 51ff.

Da die beiden ersten Terme konstant sind, interessiert nur der letzte Teil der obigen Gleichung. Für  $p_i = \frac{n_i(x)}{n}$  ergibt sich der nicht konstante Term der Likelihoodfunktion dann zu:

$$(28) \quad \sum_{i=1}^k n_i \log(n_i(x)).$$

Als Kostenfunktion wird dann definiert:

$$(29) \quad C(\text{Tab}, x) = \sum_{i=1}^k n_i \{ \log(n_i) - \log(n_i(x)) \}.$$

Eine Minimierung dieser Kosten entspricht damit einer Maximierung der Likelihoodfunktion<sup>92</sup>.

Andere Spezifikationen der Kostenfunktion, wie beispielsweise eine auf der  $\chi^2$ -Statistik basierende Spezifikation sind asymptotisch äquivalent zur oben beschriebenen Kostenfunktion. In dieser Arbeit werden Kostenfunktionen verwendet, die auf dem  $\chi^2$ -Prinzip beruhen (vgl. Paaß 1988, S. 4):

$$(30) \quad C(\text{Tab}, x) = \sum_{i=1}^k \frac{(n_i - n_i(x))^2}{n_i(x)}, \text{ für } n_i(x) > 0.$$

Bei der hier gewählten Implementation des Simulated Annealing ist an mehrere Ausgangstabellen anzupassen. Bezeichnet man mit  $\text{Tab}_i$  die zugrundegelegten Randverteilungen und mit  $x_i, i = 1, \dots, k$ , die auf der synthetischen Stichprobe basierenden und mit  $\text{Tab}_i$  korrespondierenden Tabellen, dann ergeben sich die Gesamtkosten zu:

$$(31) \quad C(\text{Tab}_1, \dots, \text{Tab}_k, x_1, \dots, x_k) = \sum_{i=1}^k C(\text{Tab}_i, x_i)$$

wobei  $C(\text{Tab}_i, x_i)$  entsprechend Gleichung (34) berechnet wird.

Die so definierte Kostenfunktion nimmt den Wert Null genau dann an, wenn eine vollständige Anpassung erreicht wird und besitzt sonst nur positive Werte. Für den Fall, daß  $n_i(x) = 0$  gilt, d.h. eine Zelle unbesetzt ist, geht der Wert  $n_i$  in die Funktion ein. Dieses Vorgehen trägt der Bedeutung der  $i$ -ten Zelle derart Rechnung, daß der Wert der Kostenfunktion dann stark vergrößert wird, wenn die Ausprägung in der Soll-Tabelle groß ist. Falls die entsprechende Zellbesetzungs-

---

<sup>92</sup>  $\sum_{i=1}^k n_i \log(n_i)$  wird in die Kostenfunktion mit aufgenommen, damit sich bei einer vollständigen Anpassung für die Kostenfunktion der Wert Null ergibt.

zahl in der Soll-Tabelle niedrig ist, dann ist die Produktion von Nullzellen in der "synthetischen" Tabelle weniger bedeutend.

### 3.3.5 Konvergenz gegen die "entropiemaximale Verteilung"

Zu den vorgegebenen Randverteilungen können mehrere Stichproben gehören, die den gleichen minimalen Kostenfunktionswert produzieren. Dies kann an folgendem kleinen Beispiel verdeutlicht werden (vgl. Tabelle 3).

Von zwei Variablen  $x_1$  und  $x_2$  mit jeweils zwei Ausprägungen (1 und 2) seien nur die eindimensionalen Randverteilungen bekannt. Gesucht ist die gemeinsame Verteilung, die eine gute Anpassung an die vorgegebenen Randbedingungen besitzt. Unter vielen anderen gemeinsamen Verteilungen sind die Verteilungen in Tabelle 3 denkbar.

Teil (a) der Tabelle 3 zeigt die gemeinsame Verteilung bei vollständiger Unabhängigkeit<sup>93</sup> der Merkmale  $x_1$  und  $x_2$  ( $\chi^2 = 0$ ); in der alternativen gemeinsamen Verteilung (Tabelle 3 (b)) sind die beiden Merkmale nicht voneinander unabhängig ( $\chi^2 = 16,67$ ). Berechnet man für beide Verteilungen den Wert der absoluten Entropie<sup>94</sup>, dann ergibt sich für Teil (a) der Tabelle 3:

$$H(p) = H(0,2;0,2;0,3;0,3) = - \sum_{i=1}^4 p_i \ln p_i = 1,37$$

und für Teil (b) der Tabelle 3:

$$H(p) = H(0,3;0,1;0,2;0,4) = - \sum_{i=1}^4 p_i \ln p_i = 1,28$$

Entsprechend den in Abschnitt III.3.2 formulierten Überlegungen weist die gemeinsame Verteilung mit der größeren Unbestimmtheit einen höheren Wert der Entropie aus, wobei in beiden Fällen der Wert der Kostenfunktion gleich Null wäre. Es ist folglich zu prüfen, ob das Simulated Annealing entropiemaximale Verteilungen bezüglich der a priori-Verteilung  $q$  erzeugt, die dann eindeutig wä-

93 Vollständige Unabhängigkeit liegt dann vor, wenn sich die Zellbesetzung durch Multiplikation der jeweiligen Randverteilung für  $x_1$  bzw.  $x_2$  dividiert durch die gesamte Fallzahl ergeben. Die Zelhäufigkeit für  $x_1 = 1$  und  $x_2 = 1$  ergibt sich folglich bei vollständiger Unabhängigkeit zu:  $(40 \times 50)/100 = 20$ .

94 Die absolute Entropie mißt den Grad der Unbestimmtheit in der Verteilung. Im Falle einer Gleichverteilung ergibt sich das Entropiemaximum zu:  $H(0,25;0,25;0,25;0,25) = 4 \times (0,25 \times \ln 0,25) = 1,3863$ . Enthält die Verteilung keine Unbestimmtheit, ergibt sich die Entropie zu:  $H(1,0,0,0) = 0$  (wegen der Definition  $0 \ln 0 := 0$ ).

ren, und wie die Konvergenz zur entropiemaximalen Verteilung von den Parameterspezifikationen abhängt.

Tabelle 3: Gemeinsame Verteilung zweier binärer Variablen bei vorgegebenen eindimensionalen Randverteilungen

(a)

x2	x1			Summe vorg.
	1	2		
	1	20	20	40
2	30	30	60	
Summe vorg.	50	50	100	

(b)

x2	x1			Summe vorg.
	1	2		
	1	30	10	40
2	20	40	60	
Summe vorg.	50	50	100	

Der folgende Satz zeigt die Konvergenz der durch das Simulated Annealing erzeugten Verteilungen  $f^n$  (mit  $n$  als Stichprobenumfang) gegen die Verteilung  $p^*$  (vgl. Abschnitt III.3.1 und III.3.2; vgl. Gyárfás/Quinke 1993, S. 157)<sup>95</sup>.

Satz 6: Sei  $a$  eine beliebige positive Zahl und  $T(n) = \frac{a}{n^2}$ . Dann gilt:

$$(32) \quad \forall \varepsilon > 0 \quad \lim_{n \rightarrow \infty} \Pr\left(\|f^n - p^*\| \leq \varepsilon\right) = 1.$$

Zum Beweis des Satzes zeigen Gyárfás/Quinke die Gültigkeit von zwei Aussagen:

95  $p^*$  ist die beste Schätzung für die unbekannte Verteilung  $p^0$ , die die relative Entropie bezüglich der a priori-Verteilung maximiert.  $f^n$  bezeichnet erzeugte Verteilungen.  $\Pr(\cdot)$  steht für die Wahrscheinlichkeit, daß sich die durch das Simulated Annealing erzeugten Verteilungen für  $n \rightarrow \infty$  der Verteilung  $p^*$  annähern.

Die detaillierte Darstellung des Beweises findet sich bei Gyárfás/Quinke 1993, S. 157ff.

1. Die Wahrscheinlichkeit, daß die durch das Simulated Annealing erzeugten Verteilungen  $f^n$  für  $n \rightarrow \infty$  in der Menge der entropiemaximalen Verteilungen enthalten ist, konvergiert unter der Bedingung, daß es sich um kostenoptimale Verteilungen handelt, gegen eins.
2. Die Wahrscheinlichkeit, kostenoptimale Verteilungen  $f^n$  zu erhalten, konvergiert für wachsenden Stichprobenumfang ( $n \rightarrow \infty$ ) gegen eins.

Wenn diese beiden Aussagen zutreffen, dann läßt sich zeigen, daß die Schnittmenge aller entropiemaximalen und kostenoptimalen Verteilungen eine Teilmenge der entropiemaximalen Verteilungen ist. Demnach sind alle kostenoptimalen Verteilungen (für  $n \rightarrow \infty$ ) auch entropiemaximale Verteilungen. Für einen hinreichend großen Stichprobenumfang ist demnach eine Annäherung der durch das Simulated Annealing erzeugten Verteilung  $f^n$  an die Verteilung  $p^*$  zu erwarten. Das Simulated Annealing erzeugt also kostenminimale Verteilungen, die gleichzeitig asymptotisch gegen die entropiemaximalen Verteilungen konvergieren.

In Satz 4 wird der Kontrollparameter in Abhängigkeit vom Stichprobenumfang definiert. Gyárfás/Quinke (1993, S. 160) zeigen, daß bei einem konstanten Kontrollparameter ( $T > 0$ ) und wachsendem Stichprobenumfang die Initialverteilung  $q$  dominiert und die Annäherung an die Vorgaben immer schlechter wird. In jeder Iteration des Algorithmus wird nur ein Fall mit Gewicht  $1/n$  in eine andere Zelle umgruppiert, so daß für  $T$  konstant die Kostenänderungen mit wachsendem  $n$  immer kleiner werden und damit die Akzeptanzwahrscheinlichkeit gegen eins konvergiert. "Da die Kostenänderungen je Iteration mindestens mit der Rate  $1/n$  und höchstens mit der Rate  $1/n^2$  gegen Null konvergieren, läßt die ... formulierte Regel für die Temperaturabsenkung ( $T = a/n^2$ ) erwarten, daß die Anzahl der Akzeptierten Verschlechterungen einen relativ stabilen (leicht sinkenden) Anteil aller Iterationen bilden wird" (Gyárfás/Quinke 1993, S.160). Bei der in dieser Arbeit gewählten Implementation des Simulated Annealing wird der Kontrollparameter experimentell (in Abhängigkeit vom Stichprobenumfang) so bestimmt, daß zunächst ca. 80% der Verschlechterungen akzeptiert werden. Durch eine entsprechende Absenkungsvorschrift reduziert sich dieser Prozentsatz nach einer großen Zahl von Iterationen bei einem konstanten Wert jeweils um ca. 5%.

### 3.3.6 Zusammenfassung der theoretischen Analyse des Simulated Annealing

Die vom Simulated Annealing erzeugte Folge von Zuständen wurde als Markov-Prozeß modelliert. Unterstellt man für eine bestimmte Zahl von Iterationen einen konstanten Wert des Kontrollparameters und reduziert man diesen Wert im Laufe des Iterationsprozesses, dann resultiert eine Folge homogener Markovketten. Zunächst wurde gezeigt, daß die Markovketten, die zu Gleichung (15) und (16) gehören, eine stationäre Verteilung besitzen. Es schließt sich ein Satz an, der zeigt,

daß die Wahrscheinlichkeit, einen global kostenminimalen Zustand zu erreichen bei einer Absenkung des Kontrollparameters nahe an Null gegen eins konvergiert. Zentral für das Erreichen eines global optimalen Zustands ist die Art und Weise, wie der Kontrollparameter abgesenkt wird. Zahlreiche Autoren formulieren Bedingungen für den Absenkungsprozeß, die eine theoretische Konvergenz sicherstellen. Die praktische Anwendbarkeit dieser sehr "elaborierten" Kühlpläne ist jedoch begrenzt, da die Temperatur zu langsam abgesenkt wird und somit unverhältnismäßig hohe Rechenzeiten die Folge wären. In der Praxis werden daher Kühlpläne gewählt, die auf heuristischen Überlegungen beruhen, das heißt die Qualität eines Kühlplans wird durch die resultierende Anpassungsgüte beurteilt.

Zu den problemspezifischen Spezifikationen gehört ebenfalls die Wahl einer Kostenfunktion, die die unterschiedlichen vom Simulated Annealing erzeugten Zustände bewertet. Für das sich hier stellende Problem wird eine dem  $\chi^2$ -Maß entsprechende Kostenfunktion vorgeschlagen, die asymptotisch äquivalent einer auf Basis des Maximum-Likelihood-Prinzips abgeleiteten Kostenfunktion ist (vgl. Paaß 1988, S. 4f, Bishop et. al. 1975, S. 513ff).

Zu jeder Stichprobe gehört eine empirische Randverteilungen, die mit den vorgegebenen Verteilungen verglichen und durch die Kostenfunktion bewertet werden. Umgekehrt entsprechen gegebenen empirischen Randverteilungen (und damit im Falle einer vollständigen Anpassung auch den vorgegebenen Verteilungen) verschiedene Stichproben. Es wurde gezeigt, daß von den potentiell möglichen Stichproben durch das Simulated Annealing diejenige erzeugt wird, die entropiemaximal gegenüber der entsprechenden a priori-Verteilung ist. Läßt man das Prinzip der "Maximalen Entropie" als Konstruktionsprinzip für eine unbekannte Verteilung  $p^*$  gelten, dann zeigt die theoretische Analyse, daß die Konstruktion synthetischer Daten mittels des Simulated Annealing eine sinnvolle Alternative zur Gewinnung von Mikrodaten darstellt.

An die theoretische Analyse schließt sich nun eine experimentelle Analyse des Verfahrens an. In einer realen Anwendung (hier: Konstruktion einer synthetischen Unternehmensstichprobe) ist die entropiemaximale Verteilung unbekannt und die Qualität der erzeugten Daten kann nur durch die Anpassung an die vorgegebenen Randverteilungen beurteilt werden. Es ist deshalb naheliegend, das Simulated Annealing für ein Problem zu implementieren, bei dem die entropiemaximale Verteilung bekannt ist. In dieser experimentellen Konstruktion kann dann der Einfluß unterschiedlicher Parameterkonstellationen auf das Erreichen der entropiemaximalen Verteilung einerseits und den kostenminimalen Zustand andererseits analysiert werden.

#### 4. Ergebnisse kleinerer Experimente

Der Algorithmus wurde in Turbo Pascal Version 7.0 programmiert. Die Auswertung der Daten erfolgte mit MS Excel Version 4.0 und SPSS. Alle Berechnungen wurden auf einem Personalcomputer mit 80486 DX-33 Prozessor durchgeführt.

Von drei binären Variablen  $x_1$ ,  $x_2$ ,  $x_3$  seien die jeweils paarweisen Randverteilungen bekannt (vgl. Tabelle 4). Gesucht ist die gemeinsame Verteilung aller Variablen. Für dieses einfache Problem läßt sich die zugehörige entropiemaximale Verteilung mittels des IPF-Algorithmus<sup>96</sup> exakt berechnen (vgl. Tabelle 5).

Tabelle 4: Vorgegebene paarweise Randverteilungen (Angaben in Prozent)

x1	x2	x3	vorgegeben	x1	x2	x3	vorgegeben	x1	x2	x3	vorgegeben
1	1		24,5	1		1	12,0		1	1	16,0
1	2		25,5	1		2	38,0		1	2	38,5
2	1		30,0	2		1	19,5		2	1	15,5
2	2		20,0	2		2	30,5		2	2	30,0
			100,0				100,0				100,0

Quelle: Gyárfás 1990, S. 80

Tabelle 5: Entropiemaximale Verteilung zu den Randverteilungen in Tabelle 4

x1	x2	x3	entr. max.	x1	x2	x3	entr. max.
1	1	1	5,18	2	1	1	10,82
1	1	2	19,32	2	1	2	19,18
1	2	1	6,82	2	2	1	8,68
1	2	2	18,68	2	2	2	11,32

Quelle: Gyárfás 1990, S. 81

<sup>96</sup> Zum IPF-Algorithmus (Iterative Proportional Fitting) vgl. Blien/Graef 1991, S. 400ff.

#### 4.1 Spezifikation der Parameter

Das Ziel des Optimierungsprozesses besteht darin, eine gemeinsame Verteilung der Variablen  $x_1, x_2, x_3$  derart zu bestimmen, daß die Abweichung von den Vorgaben möglichst gering (im Idealfall Null) ist. Die Abweichung wird durch die in Abschnitt 3.3.4 definierte Kostenfunktion gemessen:

$$(33) \quad C(\text{Tab}, x) = \sum_{i=1}^{12} \frac{(n_i - n_i(x))^2}{n_i(x)}$$

mit  $n_i$  als vorgegebenen und  $n_i(x)$  als auf Basis der synthetischen Stichprobe berechneten Randverteilungen.

Die a priori-Verteilung ist eine Gleichverteilung, das heißt potentielle Nachfolgestände werden mittels gleichverteilter Zufallszahlen erzeugt. Ausgehend von einer erzeugten Anfangsverteilung (vgl. Tabelle 6) wird zufällig ein Fall ausgewählt, und den jeweiligen Variablen werden neue Ausprägungen zugewiesen. Die entsprechende empirische Randverteilung wird mit den vorgegebenen Randbedingungen verglichen, und je nach berechneter Akzeptanzwahrscheinlichkeit wird die Veränderung akzeptiert.

Tabelle 6: Anfangsverteilung für einen Stichprobenumfang von 5.000 Fällen

x1	x2	x3	Anfangs- verteilung	x1	x2	x3	Anfangs- verteilung
1	1	1	12,50	2	1	1	13,40
1	1	2	12,00	2	1	2	12,20
1	2	1	12,70	2	2	1	11,80
1	2	2	13,00	2	2	2	12,40

Tabelle 6 zeigt, daß die Anfangsverteilung nur unwesentlich von einer Gleichverteilung der einzelnen Merkmalskombinationen abweicht. Der Maximalwert der Entropie für eine vollständige Gleichverteilung liegt bei  $\ln k = \ln 8 = 2,0794$ . Für die hier beispielhaft angegebene Anfangsverteilung liegt der Wert bei 2,0786, also nur geringfügig unter dem maximal möglichen. Sind die vorgegebenen Randverteilungen jedoch erfüllt, dann errechnet sich der Wert der Entropie aus Tabelle 5 zu 1,9836. Der Entropiewert der Anfangsverteilung nähert sich also dem Maximalwert der Entropie an, was für die Eignung des Zufallszahlengenerators zur Erzeugung gleichverteilter Zufallszahlen spricht. Die vom Simulated Annealing erzeugten Stichproben müßten, die Richtigkeit der theoretischen Aussagen zum Verhalten des Simulated Annealing vorausgesetzt, Entropiewerte liefern, die nahe bei dem Wert 1,9836 liegen.

Der Startwert des Kontrollparameters wird im Experiment alternativ auf 0,1; 1; 10, gesetzt, und im Laufe des Iterationsprozesses wird der Kontrollparameter

jeweils durch Multiplikation mit einem konstanten Faktor ( $< 1$ ) abgesenkt. Zu jedem Wert des Kontrollparameters wird jeweils eine konstante Zahl von Iterationen gerechnet, wobei die Zahl der Iterationen für einen bestimmten Kontrollparameterwert ein ganzes Vielfaches des Stichprobenumfangs ist. Beendet wird der Prozeß, wenn die Kostenfunktion den Wert Null erreicht, die Anpassung also vollständig gelingt oder nach 200-maligem Absenken des Kontrollparameters<sup>97</sup>. Die Tabellen im folgenden Abschnitt zeigen die Ergebnisse für unterschiedliche Parameterkonstellationen.

## 4.2 Ergebnisse

Die Simulationen werden für drei verschiedene Stichprobenumfänge ( $n = 100, 200, 600$ ), für drei unterschiedliche Startwerte des Kontrollparameters ( $0,1; 1; 10$ ) sowie für drei Absenkungsgeschwindigkeiten ( $0,3; 0,7; 0,95$ ) durchgeführt. Um keine vom Startwert des Zufallszahlengenerators abhängigen Ergebnisse zu erhalten, werden für jede Parameterkonstellation zehn Simulationsläufe mit unterschiedlichen Startwerten des Zufallszahlengenerators gerechnet. Für einige ausgewählte Parameterkonstellationen wird darüberhinaus der durchschnittliche Wert der Kostenfunktion, die Streuung der Kostenfunktionswerte und die durchschnittliche Akzeptanzwahrscheinlichkeit graphisch dargestellt. Tabelle 8 liegen jeweils zehn mit unterschiedlichen Startwerten des Zufallszahlengenerators gerechnete Simulationen zugrunde. Bei den ausgewiesenen Werten handelt es sich um die Mittelwerte aus diesen zehn Simulationen.

### Interpretation der Ergebnisse<sup>98</sup>

Bei allen Simulationsläufen (insgesamt 540) gelingt eine vollständige Anpassung an die vorgegebenen Randverteilungen, d.h. die Kostenfunktion hat einen Wert von Null. Die in Tabelle 5 angegebenen Anteilswerte werden zur Bestimmung der entropiemaximalen absoluten Häufigkeiten mit dem jeweiligen Stichprobenumfang multipliziert und entsprechend auf- oder abgerundet. Aufgrund dieser Rundungen können die Mittelwerte aus den jeweils zehn Simulationsläufen größere Entropiewerte aufweisen als der auf Basis von gerundeten ganzzahligen Werten errechnete Maximalwert.

Die arithmetischen Mittel für die absoluten Häufigkeiten der Merkmalskombinationen zeigen für alle Parameterkonstellationen gute Annäherungen an die entropiemaximalen Werte (vgl. Tabelle 8). Aus den arithmetischen Mitteln läßt sich keine eindeutig "überlegene" Parameterkonstellation erkennen. Verdichtet man

---

<sup>97</sup> Dies garantiert im Falle von Verschlechterungen eine Akzeptanzwahrscheinlichkeit von annähernd Null.

<sup>98</sup> Vgl. auch Abbildung 2 u. 5.

die in Tabelle 8 enthaltenen Informationen dahingehend, daß für den Startwert des Kontrollparameters nicht mehr differenziert wird, also für einen bestimmten Stichprobenumfang, eine bestimmte Lauflänge  $L_k$  und eine Absenkungsgeschwindigkeit sowie auch unabhängig von der Absenkungsgeschwindigkeit aggregiert wird, dann zeigen sich die theoretisch abgeleiteten Zusammenhänge (vgl. Tabelle 9).

Für einen gegebenen Stichprobenumfang und eine gegebene Absenkungsgeschwindigkeit wird die absolute Differenz zum maximalen Entropiewert mit einer Erhöhung der Lauflänge auf  $5n$  kleiner (Ausnahmen:  $n=100$ ,  $\alpha=0,3$ ;  $n=600$ ,  $\alpha=0,95$ ). Die absolute Differenz zum maximalen Entropiewert verringert sich bei Betrachtung der Gesamtmittelwerte, berechnet über 90 Iterationen (über alle Startwerte des Kontrollparameters und alle Absenkungsgeschwindigkeiten aggregiert) bei einer Lauflänge von  $5n$  (Ausnahme:  $n=200$ ). Mit zunehmendem Stichprobenumfang wird für den Gesamtmittelwert ( $L_k$  konstant) die Abweichung zum maximalen Entropiewert kleiner (Ausnahme:  $n=200$ ,  $L_k=5n$ ).

Betrachtet man nicht die arithmetischen Mittel, sondern die Zahl der Fälle, in denen die entropiemaximale Verteilung (absolute Häufigkeiten) exakt erreicht wird, dann erhält man weitergehende Informationen über die optimale Parameterkonstellation und über den Zusammenhang mit den theoretischen Ergebnissen (vgl. Tabelle 10). Für die verwendeten Stichprobenumfänge ( $n=100, 200, 600$ ) findet der Algorithmus unterschiedlich viele kostenminimale Lösungen. Definitionsgemäß bricht der Algorithmus mit Erreichen eines Kostenfunktionswerts von Null ab. Die Zahl der verschiedenen gefundenen Optimallösungen ist Tabelle 7 zu entnehmen. Tabelle 7 enthält zudem die zufällig zu erwartenden Fälle, in denen die Optimallösung erreicht wird; diese Zahl ergibt sich durch Division der durchgeführten Simulationen (90 für jeden Stichprobenumfang und zwei verschiedene Lauflängen) mit der Zahl der gefundenen Optimallösungen.

Tabelle 10 zeigt, daß die zufällig zu erwartende Zahl von gefundenen entropiemaximalen Verteilungen für alle Stichprobenumfänge, differenziert nach Lauflänge, vom Algorithmus um das doppelte übertroffen wird (z. B.  $n=100$ ,  $L_k=n$ : Zahl der gefundenen entropiemaximalen Verteilungen bei 90 Simulationen=30, zufällig erwartet=15). Bei einer Lauflänge von  $5n$  steigt die Wahrscheinlichkeit, die optimale Verteilung exakt zu erreichen für alle Stichprobenumfänge gegenüber einer Lauflänge in Höhe des Stichprobenumfangs (vgl. auch die Ergebnisse oben). Eine Erhöhung des Startwerts der Temperatur (Kontrollparameter) von 0,1 auf 1 führt in fast allen Fällen zu einer Verbesserung des Ergebnisses, während eine weitere Erhöhung auf einen Wert von zehn uneinheitliche Ergebnisse zeigt. Die Analysen zeigen, daß eine lange Lauflänge bei konstanter Temperatur, die komplementär zu einer niedrigen Absenkungsgeschwindigkeit ist, als entscheidender Parameter angesehen werden kann.

Die Abbildungen 3 und 6 verdeutlichen den Zusammenhang zwischen der Streuung der Kostenfunktionswerte in Abhängigkeit von den übrigen Modellparametern. Abbildung 3 zeigt, daß bei einer Absenkungsgeschwindigkeit von 0,3 das Kostenminimum von Null bereits nach sechsmaliger Absenkung der Temperatur erreicht wird. Die zugehörige Akzeptanzwahrscheinlichkeit für vorgeschlagene Verschlechterungen erreicht einen Endwert von ca. 45%. Abbildung 6 verdeutlicht für eine niedrige Absenkungsgeschwindigkeit (0,95) den Zusammenhang prägnanter:

- Die Zahl der zur Erreichung des Kostenminimums notwendigen Iterationen erhöht sich deutlich.
- Mit kleiner werdendem Kontrollparameter reduziert sich der durchschnittliche Kostenfunktionswert (mit Ausnahme bei akzeptierten Verschlechterungen) sowie die Streuung der Kostenfunktionswerte.
- Der kostenminimale Zustand wird bereits bei einer Akzeptanzwahrscheinlichkeit von ca. 70% erreicht.

Die Ergebnisse zeigen, daß die implementierte Version des Algorithmus die theoretisch abgeleiteten Ergebnisse liefert. Die Ergebnisse des Experiments entsprechen damit den theoretischen Befunden zur Konvergenz des Verfahrens. Mit einer größeren Zahl von Iterationen (z.B.  $5n$ ) erhöht sich die Wahrscheinlichkeit, die stationäre Verteilung zu erreichen. Da sowohl die theoretischen Ergebnisse als auch das Experiment auf die Überlegenheit einer längeren Lauflänge hindeuten und sich im Experiment eine Absenkungsgeschwindigkeit von 0,95 nicht als nachteilig erweist, wird zur Erzeugung der synthetischen Unternehmensstichprobe die Lauflänge auf  $5n$  und die Absenkungsgeschwindigkeit auf 0,95 gesetzt.

Da bei einer realen Anwendung die entropiemaximale Verteilung unbekannt ist, kann nicht abschließend geprüft werden, inwieweit diese Spezifikation der Parameter generell geeignet ist. Da aber die Ergebnisse der theoretischen Analyse des Verfahrens und des Experiments hinsichtlich der Spezifikation der Parameter sehr ähnlich sind, erscheint die oben dargestellte Spezifikation von Lauflänge und Absenkungsgeschwindigkeit als gerechtfertigt.

Tabelle 7: Zahl der kostenminimalen Zustände in Abhängigkeit vom Stichprobenumfang ( $L_k=n$  oder  $5n$ )\*

Stichprobenumfang	gefundene kostenminimale Lösungen	zu erwartende entropiemaximale Lösungen
$n=100$	6	15
$n=200$	9	10
$n=600$	17	5,3

\* Zahl der Simulationen für jeden Stichprobenumfang = 180

Tabelle 8: Anpassung an die entropiemaximale Verteilung für unterschiedliche Parameterkonstellationen

a)  $n=100, L_k=n$

Variable	alfa=0,3			alfa=0,7			alfa=0,95			
	$x_1$	$x_2$	$x_3$	T=0,1	T=10	T=1	T=10	T=0,1	T=1	T=10
1	1	1	5	5	5	4,5	4,9	4,9	4,6	4,6
1	1	2	19	19	19	19,5	19,1	19,1	19,4	19,4
1	2	1	7	7	7	7,5	7,1	7,1	7,4	7,4
1	2	2	19	19	19	18,5	18,9	18,9	18,6	18,6
2	1	1	11	11	11	11,5	11,2	11,2	11,4	11,4
2	1	2	19	19	19	18,5	18,9	18,9	18,6	18,6
2	2	1	9	9	9	8,5	8,9	8,9	8,6	8,6
2	2	2	11	11	11	11,5	11,1	11,1	11,4	11,4
Summe			100	100	100	100	100	100	100	100
Entropie			1,9848672	1,9848672	1,9848672	1,9839175	1,9848369	1,9847270	1,9842676	1,9842676
Differenz			0,00000000	0,00000000	0,00000000	0,00094965	0,00003031	0,00014020	0,00003031	0,00059959

b)  $n=100, L_k=5n$

Variable	alfa=0,3			alfa=0,7			alfa=0,95			
	$x_1$	$x_2$	$x_3$	T=0,1	T=10	T=1	T=10	T=0,1	T=1	T=10
1	1	1	5	5,1	4,8	5,4	4,4	4,7	5	5,3
1	1	2	19	18,9	19,2	18,6	19,6	19,3	19	18,9
1	2	1	7	6,9	7,2	6,6	7,6	7,3	7	6,7
1	2	2	19	19,1	18,8	19,4	18,4	18,7	19	19,1
2	1	1	11	10,9	11,2	10,6	11,6	11,3	11	10,9
2	1	2	19	19,1	18,8	19,4	18,4	18,7	19	19,1
2	2	1	9	9,1	8,8	9,4	8,4	8,7	9	9,3
2	2	2	11	10,9	11,2	10,6	11,6	11,3	11	10,9
Summe			100	100	100	100	100	100	100	100
Entropie			1,9848672	1,9848181	1,9847270	1,9838311	1,9834868	1,9845373	1,9848672	1,9848181
Differenz			0,00004906	0,00004906	0,00014020	0,00067091	0,00138036	0,00032987	0,00000000	0,00004906

## Fortsetzung Tabelle 8

c)  $n=200$ ,  $Lk=n$ 

Variable $x_1$ $x_2$ $x_3$	Entr. max.	alfa=0,3			alfa=0,7			alfa=0,95		
		T=0,1	T=1	T=10	T=0,1	T=1	T=10	T=0,1	T=1	T=10
1 1 1	10	9,9	9,7	9,8	10,3	10,6	10,4	9,6	9,8	9,7
1 1 2	39	39,1	39,3	39,2	38,7	38,4	38,6	39,4	39,2	39,3
1 2 1	14	14,1	14,3	14,2	13,7	13,4	13,6	14,4	14,2	14,3
1 2 2	37	36,9	36,7	36,8	37,3	37,6	37,4	36,6	36,8	36,7
2 1 1	22	22,1	22,3	22,2	21,7	21,4	21,6	22,4	22,2	22,3
2 1 2	38	37,9	37,7	37,8	38,3	38,6	38,4	37,6	37,8	37,7
2 2 1	17	16,9	16,7	16,8	17,3	17,6	17,4	16,6	16,8	16,7
2 2 2	23	23,1	23,3	23,2	22,7	22,4	22,6	23,4	23,2	23,3
Summe	200	200	200	200	200	200	200	200	200	200
Entropie	1,9834788	1,9833964	1,9831718	1,9832940	1,9836066	1,9835558	1,9836095	1,9830295	1,9832940	1,9831718
Differenz		0,0000824	0,0003070	0,0001847	-0,0001278	-0,0000770	-0,0001307	0,0004493	0,0001847	0,0003070

d)  $n=200$ ,  $Lk=5n$ 

Variable $x_1$ $x_2$ $x_3$	Entr. max.	alfa=0,3			alfa=0,7			alfa=0,95		
		T=0,1	T=1	T=10	T=0,1	T=1	T=10	T=0,1	T=1	T=10
1 1 1	10	10,3	10,5	10,5	9,7	10,5	9,8	10	10	10,1
1 1 2	39	38,7	38,5	38,5	39,3	38,5	39,2	39	39	38,9
1 2 1	14	13,7	13,5	13,5	14,3	13,5	14,2	14	14	13,9
1 2 2	37	37,3	37,5	37,5	36,7	37,5	36,8	37	37	37,1
2 1 1	22	21,7	21,5	21,5	22,3	21,5	22,2	22	22	21,9
2 1 2	38	38,3	38,5	38,5	37,7	38,5	37,8	38	38	38,1
2 2 1	17	17,3	17,5	17,5	16,7	17,5	16,8	17	17	17,1
2 2 2	23	22,7	22,5	22,5	23,3	22,5	23,2	23	23	22,9
Summe	200	200	200	200	200	200	200	200	200	200
Entropie	1,9834788	1,9836066	1,9835925	1,9835925	1,9831718	1,9835925	1,9832940	1,9834788	1,9834788	1,9835412
Differenz		-0,0001278	-0,0001138	-0,0001138	0,0003070	-0,0001138	0,0001847	0,0000000	0,0000000	-0,0000625

## Fortsetzung Tabelle 8

e)  $n=600$ ,  $Lk=n$ 

Variable			alfa=0,3			alfa=0,7			alfa=0,95		
x1	x2	x3	T=0,1	T=1	T=10	T=0,1	T=1	T=10	T=0,1	T=1	T=10
1	1	1	30,1	31,1	30,7	30,7	31	33	33	29	31,3
1	1	2	116,9	115,9	116,3	116,3	116	114	114	118	115,7
1	2	1	41,9	40,9	41,3	41,3	41	39	39	43	40,7
1	2	2	111,1	112,1	111,7	111,7	112	114	114	110	112,3
2	1	1	65,9	64,9	65,3	65,3	65	63	63	67	64,7
2	1	2	114,1	115,1	114,7	114,7	115	117	117	113	115,3
2	2	1	51,1	52,1	51,7	51,7	52	54	54	50	52,3
2	2	2	68,9	67,9	68,3	68,3	68	66	66	70	67,7
Summe			600	600	600	600	600	600	600	600	600
Entropie			1,9835018	1,9836107	1,9835936	1,9835936	1,9836097	1,9832112	1,9832112	1,9831266	1,9836060
Differenz			0,0001079	-0,0000010	0,0000161	0,0000161	0,0000000	0,0003986	0,0003986	0,0004832	0,0000037

f)  $n=600$ ,  $Lk=5n$ 

Variable			alfa=0,3			alfa=0,7			alfa=0,95		
x1	x2	x3	T=0,1	T=1	T=10	T=0,1	T=1	T=10	T=0,1	T=1	T=10
1	1	1	30,7	32,1	31,8	31,6	30,1	31,6	29,7	32	29,9
1	1	2	116,3	114,9	115,2	115,4	116,9	115,4	117,3	115	117,1
1	2	1	41,3	39,9	40,2	40,4	41,9	40,4	42,3	40	42,1
1	2	2	111,7	113,1	112,8	112,6	111,1	112,6	110,7	113	110,9
2	1	1	65,3	63,9	64,2	64,2	65,9	64,4	66,3	64	66,1
2	1	2	114,7	116,1	115,8	115,6	114,1	115,6	113,7	116	113,9
2	2	1	51,7	53,1	52,8	52,6	51,1	52,6	50,7	53	50,9
2	2	2	68,3	66,9	67,2	67,4	68,9	67,4	69,3	67	69,1
Summe			600	600	600	600	600	600	600	600	600
Entropie			1,9835936	1,9834993	1,9835558	1,9835825	1,9835018	1,9835825	1,9833964	1,9835203	1,9834535
Differenz			0,0000161	0,0001104	0,0000539	0,0000272	0,0001079	0,0000272	0,0002134	0,0000894	0,0001562

Tabelle 9: Anpassung an die entropiemaximale Verteilung: Aggregation über alle Startwerte des Kontrollparameters sowie über unterschiedliche Absenkungsgeschwindigkeiten

a)  $n=100, L_k=n$

Variable $x_1$ $x_2$ $x_3$	alfa=0,3		alfa=0,7		alfa=0,95		Gesamt	
	Entr. max.	Mittelwert	Mittelwert	Stdabw.	Mittelwert	Stdabw.	Mittelwert	Stdabw.
1 1 1	5,00	5,00	4,73	0,17	4,70	0,14	4,81	0,19
1 1 2	19,00	19,00	19,27	0,17	19,30	0,14	19,19	0,19
1 2 1	7,00	7,00	7,27	0,17	7,30	0,14	7,19	0,19
1 2 2	19,00	19,00	18,73	0,17	18,70	0,14	18,81	0,19
2 1 1	11,00	11,00	11,27	0,17	11,30	0,14	11,19	0,19
2 1 2	19,00	19,00	18,73	0,17	18,70	0,14	18,81	0,19
2 2 1	9,00	9,00	8,73	0,17	8,70	0,14	8,81	0,19
2 2 2	11,00	11,00	11,27	0,17	11,30	0,14	11,19	0,19
Summe	100	100	100	100	100	100	100	100
Entropie	1,9848672	1,9848672	1,9846094		1,9845373		1,9847431	
Differenz		0,00000000	0,00025776		0,00032987		0,00012405	

b)  $n=100, L_k=5n$

Variable $x_1$ $x_2$ $x_3$	alfa=0,3		alfa=0,7		alfa=0,95		Gesamt	
	Entr. max.	Mittelwert	Mittelwert	Stdabw.	Mittelwert	Stdabw.	Mittelwert	Stdabw.
1 1 1	5,00	5,13	4,83	0,29	5,13	0,12	5,03	0,33
1 1 2	19,00	18,87	19,17	0,29	18,87	0,12	18,97	0,33
1 2 1	7,00	6,87	7,17	0,29	6,87	0,12	6,97	0,33
1 2 2	19,00	19,13	18,83	0,29	19,13	0,12	19,03	0,33
2 1 1	11,00	10,87	11,17	0,29	10,87	0,12	10,97	0,33
2 1 2	19,00	19,13	18,83	0,29	19,13	0,12	19,03	0,33
2 2 1	9,00	9,13	8,83	0,29	9,13	0,12	9,03	0,33
2 2 2	11,00	10,87	11,17	0,29	10,87	0,12	10,97	0,33
Summe	100	100	100	100	100	100	100	100
Entropie	1,9848672	1,9847842	1,984725		1,9847842		1,9848596	
Differenz		0,00008302	0,00009472		0,00008302		0,00000734	

Fortsetzung Tabelle 9

c)  $n=200$ ,  $L_k=n$

Variable		alfa=0,3		alfa=0,7		alfa=0,95		Gesamt		
$x_1$	$x_2$	Entr. max.	Mittelwert	Stdabw.	Mittelwert	Stdabw.	Mittelwert	Stdabw.	Mittelwert	Stdabw.
1	1	1	10,00	9,80	10,43	0,12	9,70	0,08	9,98	0,34
1	1	2	39,00	39,20	38,57	0,08	39,30	0,08	39,02	0,34
1	2	1	14,00	14,20	13,57	0,08	14,30	0,08	14,02	0,34
1	2	2	37,00	36,80	37,43	0,08	36,70	0,08	36,98	0,34
2	1	1	22,00	22,20	21,57	0,12	22,30	0,08	22,02	0,34
2	1	2	38,00	37,80	38,43	0,08	37,70	0,08	37,98	0,34
2	2	1	17,00	16,80	17,43	0,12	16,70	0,08	16,98	0,34
2	2	2	23,00	23,20	22,57	0,12	23,30	0,08	23,02	0,34
Summe		200	200	200	200	200	200	200	200	200
Entropie		1,9834788	1,9832940	1,9836060	1,9831718	1,9831718	1,9831718	1,9834622	1,9834622	1,9834622
Differenz			0,0001847		-0,0001273		0,0003070		0,0000166	

d)  $n=200$ ,  $L_k=5n$

Variable		alfa=0,3		alfa=0,7		alfa=0,95		Gesamt		
$x_1$	$x_2$	Entr. max.	Mittelwert	Stdabw.	Mittelwert	Stdabw.	Mittelwert	Stdabw.	Mittelwert	Stdabw.
1	1	1	10,00	10,43	10,00	0,36	10,03	0,05	10,16	0,29
1	1	2	39,00	38,57	39,00	0,36	38,97	0,05	38,84	0,29
1	2	1	14,00	13,57	14,00	0,36	13,97	0,05	13,84	0,29
1	2	2	37,00	37,43	37,00	0,36	37,03	0,05	37,16	0,29
2	1	1	22,00	21,57	22,00	0,36	21,97	0,05	21,84	0,29
2	1	2	38,00	38,43	38,00	0,36	38,03	0,05	38,16	0,29
2	2	1	17,00	17,43	17,00	0,36	17,03	0,05	17,16	0,29
2	2	2	23,00	22,57	23,00	0,36	22,97	0,05	22,84	0,29
Summe		200	200	200	200	200	200	200	200	200
Entropie		1,9834788	1,9836060	1,9834788	1,9835018	1,9835018	1,9835018	1,9835673	1,9835673	1,9835673
Differenz			-0,0001273		0,0000000		-0,0000230		-0,0000886	

## Fortsetzung Tabelle 9

e)  $n=600$ ,  $L_k=n$ 

Variable	Entz. max.		alfa=0,3		alfa=0,7		alfa=0,95		Gesamt	
	$x_1$	$x_2$	Mittelwert	Stdabw.	Mittelwert	Stdabw.	Mittelwert	Stdabw.	Mittelwert	Stdabw.
1	1	1	31,00	0,41	31,57	1,02	31,10	1,64	31,10	1,20
1	1	2	116,00	0,41	115,43	1,02	115,90	1,64	115,90	1,20
1	2	1	41,00	0,41	41,37	1,02	40,43	1,64	40,90	1,20
1	2	2	112,00	0,41	112,57	1,02	112,10	1,64	112,10	1,20
2	1	1	65,00	0,41	64,43	1,02	64,90	1,64	64,90	1,20
2	1	2	115,00	0,41	114,63	1,02	115,57	1,64	115,10	1,20
2	2	1	52,00	0,41	51,63	1,02	52,10	1,64	52,10	1,20
2	2	2	68,00	0,41	67,43	1,02	67,90	1,64	67,90	1,20
Summe			600		600		600		600	
Entropie			1,9836097		1,9835861		1,9836107		1,9836107	
Differenz			0,0000224		0,0000237		-0,0000010		-0,0000010	

f)  $n=600$ ,  $L_k=5n$ 

Variable	Entz. max.		alfa=0,3		alfa=0,7		alfa=0,95		Gesamt	
	$x_1$	$x_2$	Mittelwert	Stdabw.	Mittelwert	Stdabw.	Mittelwert	Stdabw.	Mittelwert	Stdabw.
1	1	1	31,00	0,60	31,10	0,71	30,53	1,04	31,06	0,90
1	1	2	116,00	0,60	115,47	0,71	116,47	1,04	115,94	0,90
1	2	1	41,00	0,60	40,47	0,71	41,47	1,04	40,94	0,90
1	2	2	112,00	0,60	112,53	0,71	111,53	1,04	112,06	0,90
2	1	1	65,00	0,60	64,47	0,71	65,47	1,04	64,94	0,90
2	1	2	115,00	0,60	115,53	0,71	114,53	1,04	115,06	0,90
2	2	1	52,00	0,60	52,53	0,71	51,53	1,04	52,06	0,90
2	2	2	68,00	0,60	67,47	0,71	68,47	1,04	67,94	0,90
Summe			600		600		600		600	
Entropie			1,9835894		1,9836107		1,9835761		1,9836105	
Differenz			0,0000203		-0,0000010		0,0000336		-0,0000008	

Tabelle 10: Gefundene entropiemaximale Verteilungen für unterschiedliche Parameterkonstellationen

n=100

		Lk=n				Lk=5n			
		$\alpha = 0,3$	$\alpha = 0,7$	$\alpha = 0,95$	$\Sigma T$	$\alpha = 0,3$	$\alpha = 0,7$	$\alpha = 0,95$	$\Sigma T$
T	0,1	1	4	3	8	4	3	4	11
	1	3	3	4	10	5	3	3	11
	10	2	6	4	12	2	6	5	13
$\Sigma\alpha$		6	13	11	30	11	12	12	35

n=200

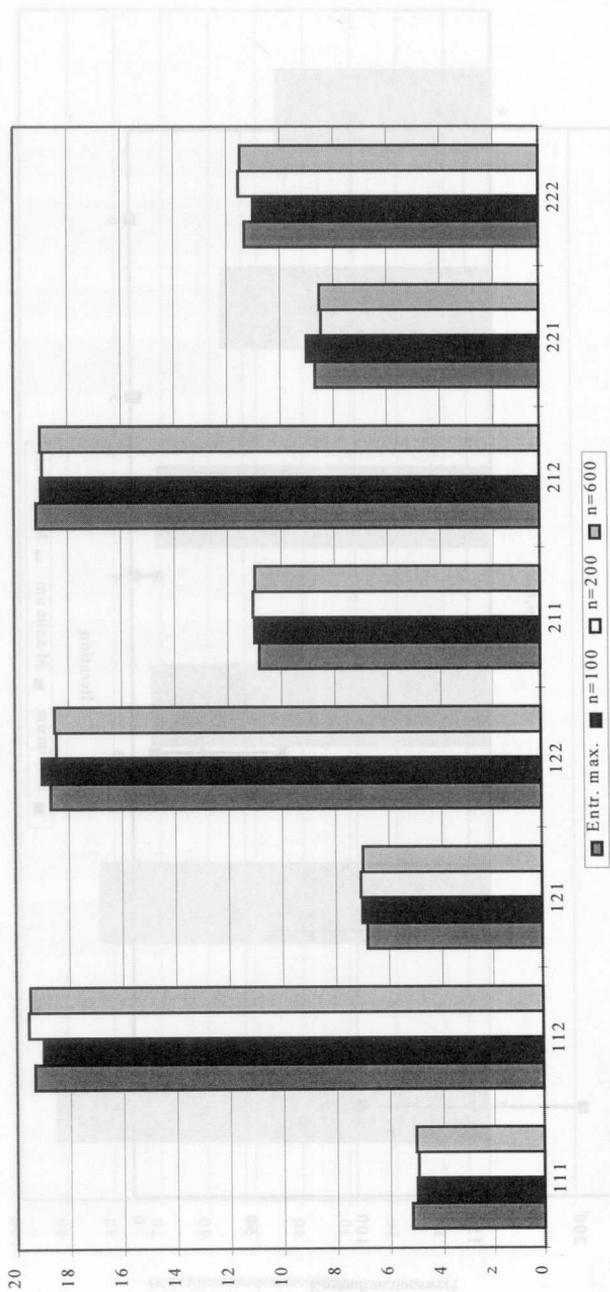
		Lk=n				Lk=5n			
		$\alpha = 0,3$	$\alpha = 0,7$	$\alpha = 0,95$	$\Sigma T$	$\alpha = 0,3$	$\alpha = 0,7$	$\alpha = 0,95$	$\Sigma T$
T	0,1	2	1	2	5	2	2	0	4
	1	1	3	3	7	5	4	5	14
	10	5	1	2	8	2	2	1	5
$\Sigma\alpha$		8	5	7	20	9	8	6	23

n=600

		Lk=n				Lk=5n			
		$\alpha = 0,3$	$\alpha = 0,7$	$\alpha = 0,95$	$\Sigma T$	$\alpha = 0,3$	$\alpha = 0,7$	$\alpha = 0,95$	$\Sigma T$
T	0,1	1	1	0	2	0	1	0	1
	1	2	1	2	5	2	1	2	5
	10	0	1	1	2	1	1	3	5
$\Sigma\alpha$		3	3	3	9	3	3	5	11

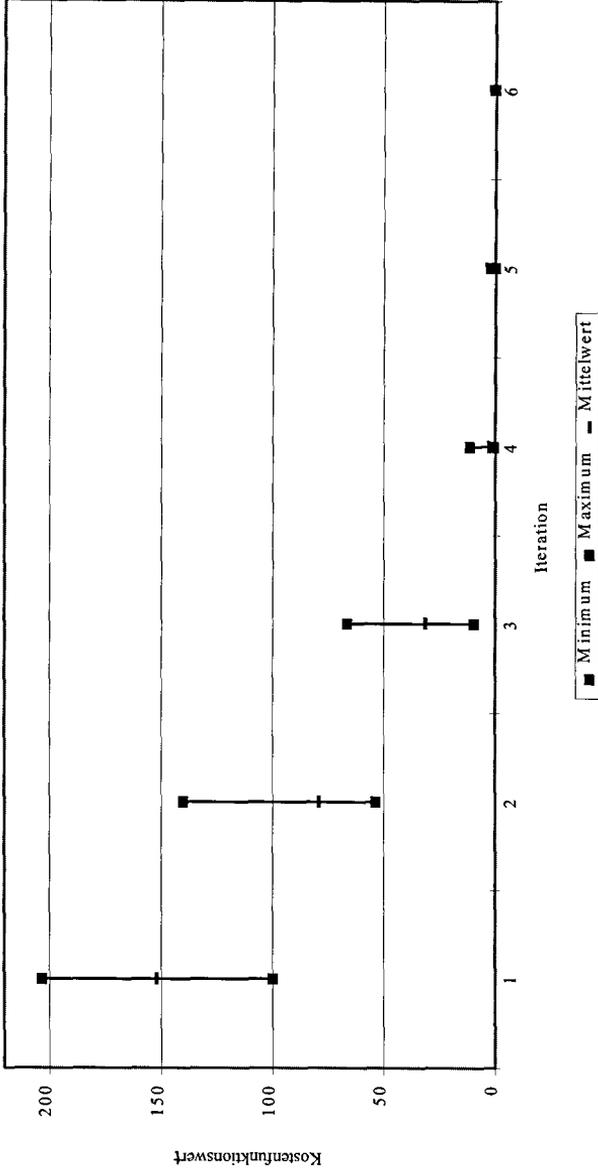
Quelle: eigene Erstellung

Abbildung 2: Anpassung an die entropiemale Verteilung in Abhängigkeit vom Stichprobenumfang für einen niedrigen Startwert des Kontrollparameters (0,1), hohe Absenkungsgeschwindigkeit (0,3) und Länge der Markovkette = n



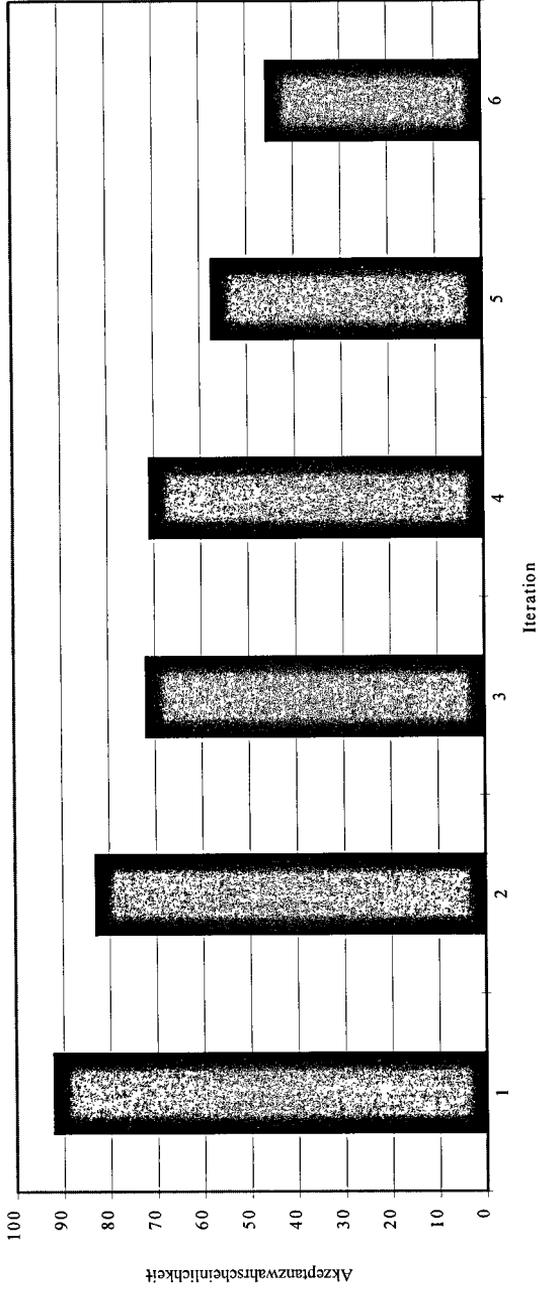
Quelle: eigene Erstellung

Abbildung 3: Mittelwert und Streuung der Kostenfunktionswerte für  $n = 600$ ;  $T_{\text{start}} = 10$ ;  $\alpha = 0,3$  und  $L_k = n$



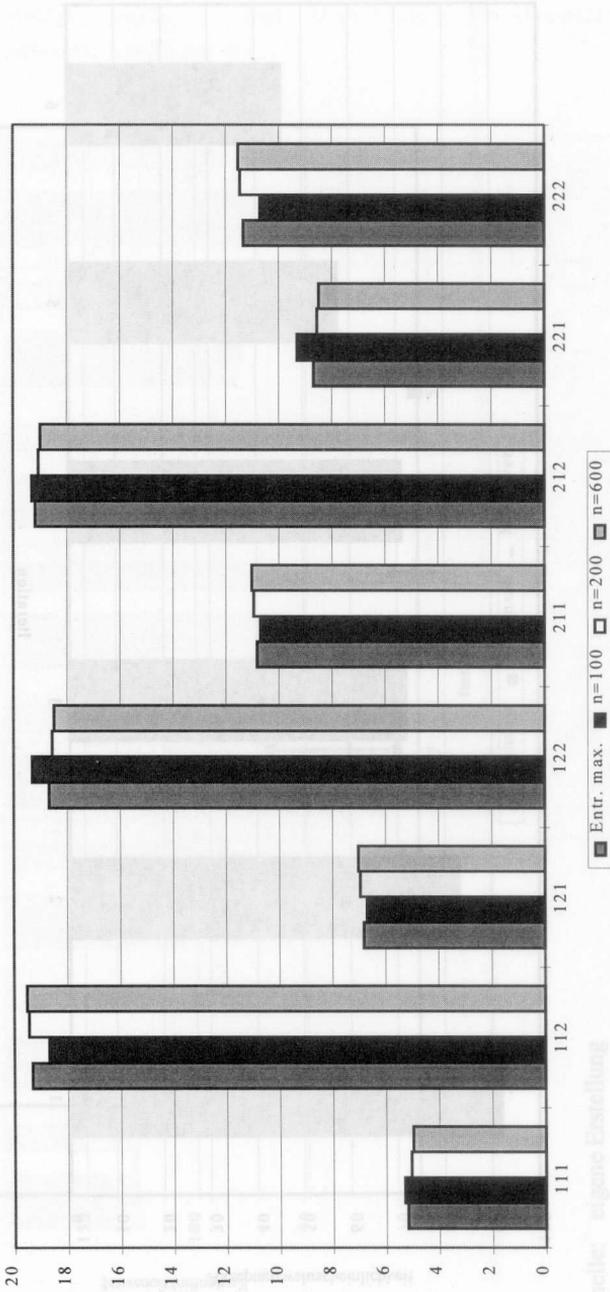
Quelle: eigene Erstellung

Abbildung 4: Durchschnittliche Akzeptanzwahrscheinlichkeit bei Verschlechterungen für  $n = 600$ ;  $T_{\text{start}} = 10$ ;  $\alpha = 0,3$  und  $L_k = n$



Quelle: eigene Erstellung

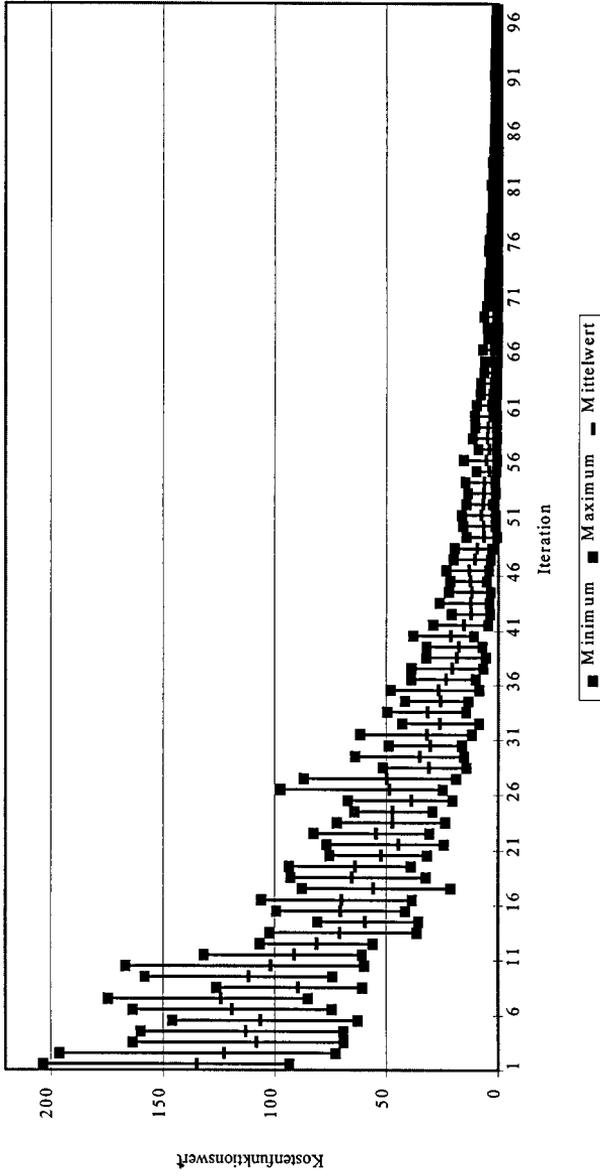
Abbildung 5: Anpassung an die entropiemale Verteilung in Abhängigkeit vom Stichprobenumfang für einen niedrigen Startwert des Kontrollparameters (10), niedrige Absenkungsgeschwindigkeit (0,95) und Länge der Markovkette  $t_e = 5n$



Quelle: eigene Erstellung

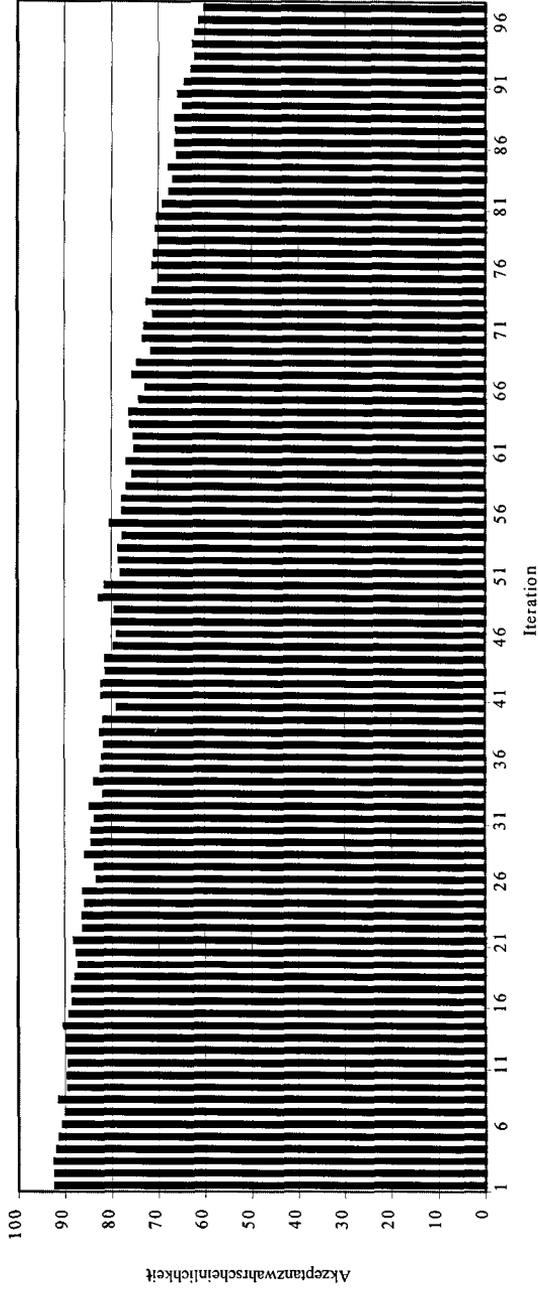
Quelle: eigene Erstellung

Abbildung 6: Mittelwerte und Streuung der Kostenfunktionswerte für  $n = 600$ ;  $T_{\text{start}} = 10$ ;  $\alpha = 0,95$  und  $L_k = 5n$



Quelle: eigene Erstellung

Abbildung 7: Durchschnittliche Akzeptanzwahrscheinlichkeit bei Verschlechterungen für  $n = 600$ ;  $T_{\text{start}} = 10$ ;  $\alpha = 0,95$  und  $L_k = 5n$



Quelle: eigene Erstellung

## 5. Rekonstruktion einer realen Stichprobe

Neben der theoretischen Analyse des Simulated Annealing und den Simulationsexperimenten im vorherigen Abschnitt wird als weiterer Test dafür, inwieweit das Simulated Annealing in der Lage ist, Stichproben zu erzeugen, die als Schätzung für die unbekannte Verteilung der Grundgesamtheit gelten können, die Rekonstruktion der Verteilung einiger Merkmale aus der "Bonner Stichprobe"<sup>99</sup> durchgeführt.

Ausgewählte Daten dieser Stichprobe für das Jahr 1986 werden in bestimmte vorgegebene Klassen aggregiert. Im einzelnen werden folgende Aggregatdaten zugrundegelegt (vgl. Tabelle 1 - 5 im Anhang): Anzahl der Unternehmen nach Umsatz- und Beschäftigtengrößenklassen, Anzahl der Unternehmen nach Lohn- und Beschäftigtengrößenklassen sowie die Lohn-, Beschäftigten- und Umsatzsumme in der jeweiligen Größenklasse. Die folgenden Tabellen 11-14 zeigen die Anpassungen der synthetischen Stichprobe an die vorgegebenen Randbedingungen einerseits und die Reproduktion von Mittelwerten und Standardabweichungen andererseits.

Tabelle 13 und 14 beinhalten die Verteilung der synthetisch erzeugten Unternehmen nach Lohn- und Beschäftigtengrößenklassen bzw. nach Umsatz- und Beschäftigtengrößenklassen. Die vorgegebenen Verteilungen werden durch die synthetisch erzeugte Stichprobe gut reproduziert; in einigen Zellen wird sogar eine exakte Abbildung der Zahl der Unternehmen erreicht. In den offenen Klassen (>25.000 Beschäftigte bzw. >1.000 Mio. Lohnsumme oder >5.000 Mio. Umsatz) wird die Zahl der Unternehmen unterschätzt. Da die Randintervalle eine wesentlich größere Spannweite aufweisen als die anderen Größenklassen (es werden für die Randintervalle gleichverteilte Zufallszahlen aus einem sehr großen Intervall gezogen), wird der Mittelwert in den Randintervallen überschätzt, so daß bei gleichzeitiger Restriktion durch die jeweilige Variablensumme, die für die drei Variablen annähernd exakt reproduziert wird, die Zahl der Unternehmen niedriger sein muß als in der Originalstichprobe (vgl. Tabelle 12).

Nicht explizit als Randbedingungen zugrundegelegt wurden der Mittelwert, die Standardabweichung sowie das Minimum und Maximum der Variablen. Dennoch werden diese Größen im Vergleich zur Originalstichprobe gut abgebildet, wenn auch in den Randintervallen größere Abweichungen zu beobachten sind (vgl. Tabelle 12). Bei der Analyse der Korrelationskoeffizienten (vgl. Tabelle 11) zeigt

---

<sup>99</sup> Die Bonner Stichprobe enthält die veröffentlichten Bilanzdaten deutscher Industrieaktiengesellschaften. Insgesamt sind 717 Unternehmen seit 1960 erfaßt (Stand 1990), für die jedoch nicht immer die gesamte Zeitreihe zur Verfügung steht.

sich, daß die Korrelation zwischen den Variablen in der synthetischen Stichprobe zwar die gleiche Richtung aufweist, betragsmäßig aber kleiner ist.

Tabelle 11: Vergleich der Korrelationskoeffizienten in der Bonner Stichprobe und der rekonstruierten Stichprobe\*

BS: Bonner Stichprobe, SS: Synthetische Stichprobe

	Lohn		Umsatz		Beschäftigte	
	BS	SS	BS	SS	BS	SS
Lohn	1,00	1,00	0,95	0,51	0,93	0,84
Umsatz	0,95	0,51	1,00	1,00	0,99	0,61
Beschäftigte	0,93	0,84	0,99	0,61	1,00	1,00

\* alle Koeffizienten sind signifikant auf dem 0,001-Niveau

Quelle: eigene Erstellung

Aufgrund der Unterschätzung der betragsmäßigen Korrelation zwischen den einzelnen Variablen ist zu erwarten, daß auf Basis der synthetischen Stichprobe gefundene Zusammenhänge, die korrekte Richtung des Wirkungszusammenhangs widerspiegeln, der quantitative Einfluß jedoch unterschätzt wird. Durch das verwendete Konstruktionsprinzip der maximalen Entropie wird also ein eher vorsichtiger Weg beschritten, d.h. auf Basis "echter" Daten könnten die Zusammenhänge deutlicher hervortreten.



## Fortsetzung Tabelle 12:

## b) Variable Umsatz

Größenklasse	Summe			Mittelwert			Standardabweichung		
	BS	SS	Differenz	BS	SS	Differenz	BS	SS	Differenz
unter 10 Mio.	69	70	-1,08	3,30	2,60	21,21	2,70	2,30	14,81
10 b. u. 50 Mio.	1.315	1.314	0,08	29,20	27,40	6,16	10,90	12,30	-12,84
50 b. u. 100 Mio.	3.357	3.353	0,11	74,50	72,90	2,15	13,80	15,80	-14,49
100 b. u. 200 Mio.	8.197	8.194	0,04	134,30	134,30	0,00	26,80	23,20	13,43
200 b. u. 500 Mio.	20.028	20.028	0,00	323,00	323,00	0,00	86,00	80,40	6,51
500 b. u. 1.000 Mio.	30.820	30.840	-0,06	716,70	700,90	2,20	131,40	143,60	-9,28
1.000 b. u. 5.000 Mio.	142.910	143.143	-0,16	2.342,70	2.468,00	-5,35	1.104,30	1.128,20	-2,16
5.000 Mio. u. mehr	769.902	771.812	-0,25	19.741,00	24.897,20	-26,12	15.100,40	15.952,10	-5,64

Größenklasse	Minimum			Maximum		
	BS	SS	Differenz	BS	SS	Differenz
unter 10 Mio.	0,37	0,11	70,27	9,20	9,60	-4,35
10 b. u. 50 Mio.	10,40	10,20	1,92	48,50	49,80	-2,68
50 b. u. 100 Mio.	51,20	50,40	1,56	98,90	99,40	-0,51
100 b. u. 200 Mio.	100,00	101,90	-1,90	193,40	184,70	4,50
200 b. u. 500 Mio.	200,40	202,30	-0,95	496,60	497,30	-0,14
500 b. u. 1.000 Mio.	505,90	503,60	0,45	962,40	988,50	-2,71
1.000 b. u. 5.000 Mio.	1.048,80	1.020,90	2,66	4.968,50	4.865,40	2,08
5.000 Mio. u. mehr	5.136,40	5.570,40	-8,45	65.498,10	61.260,40	6,47

## Fortsetzung Tabelle 12:

## c) Variable Lohn

Größenklasse	Summe			Mittelwert			Standardabweichung		
	BS	SS	Differenz	BS	SS	Differenz	BS	SS	Differenz
unter 1 Mio.	8,83	8	9,40	0,50	0,40	20,00	0,30	0,30	0,00
1 b. u. 5 Mio.	56,25	56	0,44	2,80	2,60	7,14	1,20	1,10	8,33
5 b. u. 10 Mio.	207,60	223	-7,42	7,40	7,40	0,00	1,50	1,50	0,00
10 b. u. 20 Mio.	587,87	622	-5,81	15,10	14,80	1,99	2,50	3,00	-20,00
20 b. u. 30 Mio.	983,19	996	-1,30	24,60	24,30	1,22	2,70	2,90	-7,41
30 b. u. 50 Mio.	1.710,02	1.745	-2,05	38,90	39,70	-2,06	6,00	5,00	16,67
50 b. u. 100 Mio.	3.722,29	3.724	-0,05	74,40	73,00	1,88	16,10	14,40	10,56
100 b. u. 200 Mio.	4.567,16	4.745	-3,89	138,40	143,80	-3,90	31,60	32,40	-2,53
200 b. u. 500 Mio.	15.739,88	17.346	-10,20	321,20	346,90	-8,00	91,10	84,90	6,81
500 b. u. 1.000 Mio.	17.291,61	17.131	0,93	720,40	744,90	-3,40	137,00	112,60	17,81
1.000 Mio. u. mehr	156.927,05	157.464	-0,34	4.755,40	7.157,50	-50,51	4.063,20	4.372,90	-7,62

Größenklasse	Minimum			Maximum		
	BS	SS	Differenz	BS	SS	Differenz
unter 1 Mio.	0,03	0,02	33,33	0,90	0,99	-10,00
1 b. u. 5 Mio.	1,10	1,10	0,00	4,80	4,80	0,00
5 b. u. 10 Mio.	5,10	5,10	0,00	9,90	9,90	0,00
10 b. u. 20 Mio.	11,00	10,10	8,18	19,90	19,80	0,50
20 b. u. 30 Mio.	20,70	20,00	3,38	29,80	29,40	1,34
30 b. u. 50 Mio.	30,40	31,20	-2,63	49,40	49,20	0,40
50 b. u. 100 Mio.	50,70	50,40	0,59	99,80	96,40	3,41
100 b. u. 200 Mio.	102,30	101,30	0,98	198,40	198,40	0,00
200 b. u. 500 Mio.	202,00	202,20	-0,10	489,40	497,10	-1,57
500 b. u. 1.000 Mio.	530,30	547,40	-3,22	983,90	991,80	-0,80
1.000 Mio. u. mehr	1.118,50	1.632,80	-45,98	16.981,10	15.067,00	11,27

Tabelle 13: Anzahl der Betriebe nach Lohn- und Beschäftigengrößenklassen

Größenklasse Besch.	Größenklasse Lohn (in Millionen DM)																								Summe	
	<1		<5		<10		<20		<30		<50		<100		<200		<500		<1.000		>1.000		BGK			
	BS	SS	BS	SS	BS	SS	BS	SS	BS	SS	BS	SS	BS	SS	BS	SS	BS	SS	BS	SS	BS	SS	BS	SS		
unter 100	17	17	15	15	1																					
100 b. u. 500	3	4	4	28	28	38	38	18	18	1																
500 b. u. 1.000			1	1	1	1	21	21	32	32	1	1														
1.000 b. u. 2.000			1	1	1	1	1	1	12	12	38	38	3	3												
2.000 b. u. 5.000											11	11	27	27	11	11										
5.000 b. u. 10.000													3	3	30	30	2	2								
10.000 b. u. 25.000																										
25.000 u. mehr																										
Summe Lohngrößenkl.	17	20	20	21	28	30	39	42	40	41	44	44	50	51	33	33	49	50	24	23	33	22	377	377		

Tabelle 14: Anzahl der Betriebe nach Umsatz- und Beschäftigengrößenklassen

Größenklasse Besch.	Größenklasse Umsatz (in Millionen DM)																								Summe	
	<10		<50		<100		<200		<500		<1.000		<5.000		>5.000								BGK			
	BS	SS	BS	SS	BS	SS	BS	SS	BS	SS	BS	SS	BS	SS	BS	SS	BS	SS	BS	SS	BS	SS	BS	SS		
unter 100	19	19	10	12	1	2																				
100 b. u. 500	1	5	34	34	28	28	18	18	6	6	1	1														
500 b. u. 1.000			2	1	2	12	28	28	10	10	3	3														
1.000 b. u. 2.000			1	1	4	4	13	13	25	25	9	9	4	4												
2.000 b. u. 5.000							2	2	18	18	18	18	10	10	1	1										
5.000 b. u. 10.000									2	2	12	12	18	18	3	3										
10.000 b. u. 25.000																										
25.000 u. mehr																										
Summe Umsatzgrößenkl.	21	27	45	48	45	46	61	61	62	62	43	44	61	58	39	31	377	377								

## 6. Zusammenfassung der Simulationsergebnisse

Die Experimente (Rekonstruktion einer bekannten entropiemaximalen Verteilung, Rekonstruktion einer realen Stichprobe) zeigen insgesamt eine sehr gute Anpassung an die vorgegebenen Randverteilungen sowie bei bekannter entropiemaximaler Verteilung eine gute Reproduktion dieser Verteilung. Dabei erhöht sich die Wahrscheinlichkeit, die entropiemaximale Verteilung unter allen kostenminimalen Verteilungen zu erreichen, mit einer Erhöhung der Lauflänge bei konstanter Temperatur und niedriger Absenkungsgeschwindigkeit. Der Startwert des Kontrollparameters sollte so gewählt werden, daß zunächst etwa 80% der vorgeschlagenen Verschlechterungen akzeptiert werden. Eine darüber hinausgehende Erhöhung des Startwerts des Kontrollparameters (Temperatur) führt zu keiner weiteren Ergebnisverbesserung.

Der Vergleich zwischen der rekonstruierten und der realen Stichprobe zeigt, daß die Anpassung an die Randbedingungen auch hier sehr gut gelingt. Die Reproduktion der Stichprobenmittelwerte und -standardabweichungen erscheint ebenfalls als zufriedenstellend, wenn auch in den Randklassen größere Abweichungen festzustellen sind, die jedoch auf die bereits oben genannten Gründe zurückzuführen sind.

Insgesamt deuten die Simulationsergebnisse darauf hin, daß der hier beschrittene Weg - Konstruktion einer synthetischen Unternehmensstichprobe auf Basis verschiedener Aggregatinformationen - eine vielversprechende und kostengünstige Alternative zur Gewinnung von Unternehmensdaten darstellt. Gefundene Zusammenhänge werden eher vorsichtig abgebildet. Da aber mit großen Stichprobenumfängen gearbeitet werden kann, können signifikante Zusammenhänge eher gefunden werden als bei Analysen auf Basis von Aggregatdaten, wenn auch bei letzteren die Zusammenhänge aufgrund der Aggregation zwischen den Variablen möglicherweise wesentlich größer sind.

Die in der synthetischen Stichprobe abgebildeten Strukturen der Grundgesamtheit basieren ausschließlich auf den in den Aggregatdaten enthaltenen Strukturinformationen. Es könnte der Vorwurf erhoben werden, daß die synthetisch erzeugten Strukturen beliebigen Charakter besitzen. Einerseits zeigt die Begründung des Konstruktionsprinzips der maximalen Entropie, daß es sich bei diesem Prinzip um eine rational zu begründende Möglichkeit der Schätzung einer unbekanntenen Verteilung handelt, die von Gyárfás (1990, S. 56) als die "Leitvorstellung zur Konstruktion synthetischer Mikrodaten überhaupt" bezeichnet wird. Andererseits wird bei der Hochrechnung von realen Stichproben zur Herstellung der Repräsentativität der realisierten Stichprobe an Aggregatinformationen angepaßt. Es werden Gewichtungsfaktoren bestimmt, die eine gute Anpassung an die Aggregatdaten sicherstellen sollen, d.h. in der realisierten Stichprobe enthaltene Strukturin-

formationen werden durch Gewichtung gegebenenfalls verfälscht (vgl. hierzu den folgenden Abschnitt III.7).

## 7. Synthetische Daten und Gewichtung von Stichproben

Auch bei originär erhobenen Stichproben wird in der Regel eine Gewichtung oder Hochrechnung notwendig, da in den wenigsten Fällen eine repräsentative Stichprobe realisiert werden kann<sup>100</sup>. "Dies gilt bereits nicht bei klassischen einfachen Zufallsstichproben, bei denen derartige Anteile bzw. Mittelwerte für die entsprechenden Populationsparameter zwar erwartungstreue Schätzer sind, jedoch in Abhängigkeit vom Umfang der Stichprobe mehr oder weniger um den wahren Populationswert streuen. Bei komplexeren Stichprobenplänen ... ist die Wahrscheinlichkeit, in die Stichprobe zu gelangen, nicht - wie bei einer einfachen Zufallsstichprobe - für jedes Individuum der Grundgesamtheit gleich groß. Eine einfache Hochrechnung von Populationsanteilen oder Mittelwerten führt in einem solchen Fall in der Regel zu verzerrten Ergebnissen" (Rothe 1990, S. 31). Im Falle einer freien Hochrechnung, bei der nur mit dem reziproken Wert des Auswahlwahrscheinlichkeit hochgerechnet wird, verändert sich die Struktur der Stichprobe nicht. Bei einer gebundenen Hochrechnung werden zusätzliche Informationen über die Grundgesamtheit derart verwendet, daß die Stichprobe an Total- oder Randwerte, die aus anderen Statistiken bekannt sind, angepaßt wird. Da aber nicht alle Variablen in die Berechnung dieser Hochrechnungs- oder Gewichtungsfaktoren eingehen, muß überprüft werden, welche Anpassungen sich für nicht in die Adaption der Hochrechnungsfaktoren einbezogene Variablen ergeben.

Rothe (1990) zeigt anhand des ALLBUS 1986 eindrucksvoll, daß die Effekte der gewählten Gewichtungsprozedur auf die Hochrechnung sehr heterogen und damit entmutigend sind. Selbst bei den Variablen, die zur Berechnung der Gewichtungsfaktoren herangezogen wurden, verschlechtert sich zum Teil die Anpassung gegenüber einer einfachen Hochrechnung. Für nicht in die Berechnung der Gewichtungsfaktoren einbezogene Variablen kann er zum Teil erhebliche Verschlechterungen gegenüber einer ungewichteten Hochrechnung nachweisen.

Um möglichst viel von der in der Stichprobe enthaltenen Strukturinformation zu erhalten, sollten die neuen Hochrechnungsfaktoren sich möglichst wenig von bereits vorhandenen (gebundenen oder ungebundenen) unterscheiden. Gesucht ist demnach eine Zielfunktion für die Bestimmung optimaler Gewichtungs- oder Hochrechnungsfaktoren, die einerseits den Abstand minimiert und andererseits theoretisch begründbar ist. Merz (1983, S. 346ff) schlägt als Zielfunktion die

---

<sup>100</sup> Das Erhebungsdesign ist möglicherweise noch so ausgelegt, daß jedes Unternehmen mit gleicher Wahrscheinlichkeit ausgewählt wird. Antwortverweigerungen und/oder Inkonsistenzen führen aber häufig zu verzerrten Stichproben, die dann gewichtet werden müssen.

"Minimierung des Informationsverlustes" vor, ein Maß, das eng mit dem Prinzip der "maximalen relativen Entropie" verbunden ist.

Ausgangspunkt sei wiederum die Verteilung:

$$p = (p_1, \dots, p_n), \quad p_j > 0, \quad j = 1, \dots, n, \quad \sum_{j=1}^n p_j = 1$$

mit der Entropie oder Information:

$$(34) \quad H(p) = H(p_1, \dots, p_n) = \sum_{j=1}^n p_j \ln \frac{1}{p_j} = - \sum_{j=1}^n p_j \ln p_j$$

Der Informationsgewinn bei der Ersetzung der Verteilung  $q = (q_1, \dots, q_n)$  durch die Verteilung  $p = (p_1, \dots, p_n)$  ist definiert als:

$$(35) \quad I(p:q) = \sum_{j=1}^n p_j \ln \frac{1}{q_j} - \sum_{j=1}^n p_j \ln \frac{1}{p_j} = \sum_{j=1}^n p_j \ln \frac{p_j}{q_j}$$

$$\text{mit } (p_j, q_j > 0), \quad (j = 1, \dots, n), \quad \sum_{j=1}^n p_j = \sum_{j=1}^n q_j = 1$$

(vgl. Merz 1983, S. 347, Wauschkuhn 1982, S. 41f.). Gleichung (35) ist bis auf das Vorzeichen identisch mit der Definition der relativen Entropie. Eine Maximierung der relativen Entropie ist demnach äquivalent mit einer Minimierung des Informationsverlusts, d.h. in beiden Fällen soll die Abweichung von der Ausgangsverteilung nur so groß werden, wie die Randbedingungen dies erfordern. Bei der Anwendung des Simulated Annealing entfernt man sich im Laufe des Iterationsprozesses von der a priori-Verteilung, und im Falle der Bestimmung optimaler Gewichtungsfaktoren nimmt der Abstand zu den "alten" Gewichtungsfaktoren zu. Der Zusammenhang von Informationsverlust und Entropie läßt sich verdeutlichen, wenn für die Verteilung  $q$  eine Gleichverteilung ( $q_i = 1/n$ ) unterstellt wird. Der Informationsgewinn ergibt sich dann durch Einsetzen in Gleichung (35) zu:

$$I(p:q) = \sum_{j=1}^n p_j \ln np_j = \sum_{j=1}^n p_j \ln n - \sum_{j=1}^n p_j \ln \frac{1}{p_j} = \ln n - H(p).^{101}$$

Ist die Ausgangsverteilung eine Gleichverteilung, dann entspricht die Minimierung des Informationsverlusts einer Maximierung der Entropie. Die Funktion  $I(p:q)$  erreicht ihr Minimum (Null), wenn die Verteilungen  $p$  und  $q$  identisch sind. Dann gilt:

---

<sup>101</sup> Vgl. Wauschkuhn 1982, S. 123ff.

$$I(p:q) = \sum_{j=1}^n p_j \ln \frac{p_j}{q_j} = \sum_{j=1}^n p_j \ln 1 = 0.$$

Bei der gewichteten Hochrechnung, die wegen fast nie exakt zu realisierender Erhebungsdesigns notwendig wird, entfernt man sich unter Umständen von den wahren Strukturen in der Grundgesamtheit, wohingegen bei der Konstruktion synthetischer Mikrodaten eine Annäherung an die wahre Struktur in der Grundgesamtheit angestrebt wird. Möglicherweise ist der Unterschied in der Qualität der Abbildung von Realität durch die beiden Verfahren - Hochrechnung originärer Stichproben und Erzeugung synthetischer Stichproben - nicht besonders groß, da zum einen bei der optimalen Umgewichtung von Stichproben Strukturinformationen verloren gehen, bzw. für nicht in die Gewichtungszurückführung einbezogene Variable erhebliche Verzerrungen zu beobachten sind, und zum anderen beiden Vorgehensweisen sehr verwandte Konstruktionsprinzipien zugrunde liegen. Zudem werden bei beiden Vorgehensweisen die gleichen Randinformationen genutzt.

## **IV. Konstruktion einer synthetischen Unternehmensstichprobe zur Bestimmung der Determinanten der betrieblichen Schwerbehindertenbeschäftigung und deren Einfluß auf die Wertschöpfung im Unternehmen**

### **1. Vorbemerkung**

Nachdem in Kapitel III das Simulated Annealing detailliert analysiert wurde, die theoretischen Ergebnisse und umfangreiche Simulationsexperimente auf eine Eignung des Verfahrens zur Erzeugung synthetischer Stichproben hindeuten, wird in diesem Kapitel die Vorgehensweise bei der Konstruktion einer synthetischen Unternehmensstichprobe zur Beantwortung einer substanzwissenschaftlichen Fragestellung vorgestellt. Eine Analyse der Determinanten der betrieblichen Schwerbehindertenbeschäftigung bietet sich hierfür aus mehreren Gründen an:

- Mit den Arbeiten von Frick (1992) und Sadowski/Frick (1992) liegen umfangreiche Dokumentationen der bisherigen Forschungsergebnisse vor.
- Die von den Autoren selbst produzierten empirischen Ergebnisse basieren zum einen auf Individualdaten und zum anderen auf Aggregatdaten, so daß ein Vergleich mit Ergebnissen auf Unternehmensebene nahe liegt<sup>102</sup>.
- Die von den Autoren getesteten Hypothesen sind aus der Theorie interner Arbeitsmärkte abgeleitet, so daß sich dieses theoretische Konzept zur Herleitung der zu konstruierenden Variablen heranziehen läßt.

In den nachfolgenden Abschnitten wird daher zunächst das den Arbeiten von Frick bzw. Sadowski/Frick zugrundeliegende theoretische Konzept vorgestellt. Hieran schließt sich eine Darstellung und Diskussion der dort gewählten methodischen Vorgehensweise und eine Zusammenfassung der wichtigsten Ergebnisse an. Ausgehend von dem gewählten theoretischen Konzept werden dann die zu erzeugenden Variablen bestimmt und mittels des Simulated Annealing erzeugt. Abschließend wird die synthetische Stichprobe auf ihre "Qualität" hin untersucht und einige der von Frick (1992) und Sadowski/Frick (1992) aufgestellten Hypothesen auf Basis der synthetischen Stichprobe empirisch getestet.

---

<sup>102</sup> Frick (1992) verwendet zudem noch Betriebsdaten zur Analyse des Einflusses betrieblicher Interessenvertretungen auf die Schwerbehindertenquote.

## 2. Theoretische Analyse der betrieblichen Nachfrage nach schwerbehinderten Arbeitnehmern<sup>103</sup>

Mit dem 1974 verabschiedeten und 1986 novellierten Schwerbehindertengesetz wurden alle öffentlichen und privaten Arbeitgeber mit mehr als fünfzehn Beschäftigten verpflichtet, die berufliche (Wieder-)Eingliederung Schwerbehinderter zu fördern. Dies kann entweder durch eine unmittelbare Förderung geschehen (Besetzung von mindestens 6% der Arbeitsplätze mit Schwerbehinderten, Gleichgestellten oder sonstigen anrechnungsfähigen Personen) oder mittelbar durch Zahlung einer Ausgleichsabgabe von derzeit 200,- DM pro unbesetztem Pflichtplatz<sup>104</sup>.

Trotz gesetzlicher Normierung und Sanktionierung kommen nur 22% der beschäftigungspflichtigen Arbeitgeber ihrer Beschäftigungspflicht in vollem Umfang nach, ein Drittel der beschäftigungspflichtigen Arbeitgeber beschäftigt keinen einzigen Schwerbehinderten. "Der Anteil der Arbeitgeber, der seiner Beschäftigungspflicht nachkommt bzw. nicht (vollständig) nachkommt, ist im Zeitablauf relativ konstant geblieben. Der von 1975-1982 zu verzeichnende Anstieg der durchschnittlichen Istquote von 3,8% auf 5,9% ist somit überwiegend durch die Zunahme der Schwerbehindertenbeschäftigung in Betrieben und Dienststellen zu erklären, die ihre Pflichtquote ohnehin bereits erfüllten" (Frick 1992, S. 20). Ritz (1984, S. 18) spricht in diesem Zusammenhang von einer gewissen "Polarisierung" zwischen Betrieben bzw. Dienststellen, die die gesetzliche Pflichtquote übererfüllen, und Betrieben und Dienststellen, die die Beschäftigung Schwerbehinderter zu umgehen versuchen.

Auf Basis der Daten aus dem jährlichen Anzeigeverfahren gemäß § 13 Abs. 2 SchwbG (Schwerbehindertengesetz) läßt sich zeigen, daß die betrieblichen Schwerbehindertenquoten zwischen den in der amtlichen Statistik ausgewiesenen neun Betriebsgrößenklassen und zehn Wirtschaftsabteilungen stark streuen (vgl. Tabelle 15). "Diese Streuung der Istquote und deren relative Konstanz in den Jahren 1975-1987 sind insofern erklärungsbedürftig, als ausnahmslos alle Arbeitgeber mit mehr als fünfzehn Beschäftigten unabhängig von ihrer Branchenzugehörigkeit zur Förderung der beruflichen (Re-)Integration Schwerbehinderter verpflichtet sind" (Frick 1992, S. 20f).

---

<sup>103</sup> Vgl. zu diesem Abschnitt Frick 1992, S. 19ff; Sadowski/Frick 1992, S. 25ff; Frick 1991, S. 310ff.

<sup>104</sup> Die Zahlung einer Ausgleichsabgabe hebt die Pflicht zur Beschäftigung Schwerbehinderter jedoch nicht auf (vgl. § 11 Abs. 1 SchwbG).

Tabelle 15: Schwerbehindertenquote nach Wirtschaftsabteilungen und Betriebsgrößenklassen

Branche	Jahr					
	1982	1983	1984	1985	1986	1987
Land- und Forstwirtschaft	3,9	4,0	3,5	3,5	3,7	3,4
Energie und Bergbau	11,0	10,7	9,4	9,0	9,1	8,3
Verarbeitendes Gewerbe	6,3	6,0	5,5	5,2	5,3	5,1
Baugewerbe	3,5	3,4	3,2	3,0	3,2	3,2
Handel	4,2	4,1	3,8	3,6	3,8	3,5
Verkehr und Nachrichten	5,4	5,3	5,2	5,0	5,0	5,1
Banken und Versicherungen	4,8	4,5	4,3	4,0	4,3	4,1
Dienstleistungen	4,9	4,9	4,6	4,3	4,6	4,5
Org. ohne Erwerbscharakter	6,3	6,2	5,8	5,4	5,4	5,4
Öffentlicher Dienst	7,3	7,2	6,8	6,5	6,7	6,3
Größenklasse (Beschäftigte)						
16 bis unter 30	3,5	3,3	3,1	3,0	3,4	3,3
30 bis unter 100	4,3	4,1	3,9	3,7	4,0	3,9
100 bis unter 300	5,4	5,2	4,8	4,5	4,8	4,6
300 bis unter 500	5,8	5,6	5,2	4,9	5,2	5,0
500 bis unter 1.000	6,3	6,2	5,7	5,3	5,4	5,2
1.000 bis unter 10.000	6,8	6,5	6,1	5,6	5,8	5,7
10.000 bis unter 50.000	6,9	6,6	5,8	5,7	5,6	5,4
50.000 bis unter 100.000	4,9	5,6	5,7	5,1	5,6	5,4
100.000 und mehr	7,0	6,6	6,3	5,9	6,0	5,7
insgesamt	5,9	5,7	5,3	5,0	5,2	5,0

Quelle: Frick 1992, S. 253

Frick greift in seiner Analyse der Determinanten der betrieblichen Schwerbehindertenbeschäftigung auf das Ende der sechziger und Anfang der siebziger Jahre und inzwischen vielfach modifizierte "Konzept des internen Arbeitsmarktes" zurück. "Ein wesentliches Charakteristikum dieses Ansatzes besteht in der Annahme, daß die Verbreitung aufeinander abgestimmter Rekrutierungs-, Qualifizierungs- und Entlohnungspraktiken - die ihrerseits den internen Arbeitsmarkt konstituieren - mit spezifischen betrieblichen Merkmalen, wie z. B. der Beschäftigtenzahl, der Kapitalintensität, der Konjunkturreakibilität und der Stellung auf den relevanten Absatz- und Beschaffungsmärkten, kovariiert. Diese Annahme impliziert, daß zumindest größere Betriebe keineswegs gezwungen sind, mittels adaptiver Verhaltensweisen auf sich ändernde Umweltbedingungen zu reagieren, sondern als "strategiefähig" zu gelten haben. Damit rückt die von der neoklassischen

Theorie weitgehend vernachlässigte "Binnenstruktur" der Betriebe in das Zentrum der Analyse" (Frick 1992, S. 25).

## 2.1 Betriebe als Akteure

In der neoklassischen Theorie wird das Unternehmen als eine "black-box" betrachtet, während in der neueren Organisationstheorie Unternehmen als quasi autonome Akteure angesehen werden, die verschiedene Ressourcen bei unvollständiger Information allozieren, nutzen und transformieren. Diese Sichtweise weicht von traditionellen Sichtweisen ab, weil Unternehmen nicht notwendigerweise zur Anpassung an exogene Faktoren gezwungen sind, um im Wettbewerb zu bestehen. Vielmehr sind die Unternehmen dazu in der Lage, Autonomiestrategien zu entwickeln, die es ermöglichen, zumindest einen Teil der externen Faktoren zu kontrollieren bzw. zu verstetigen. Für den Faktormarkt "Arbeit" bedeutet dies, daß der Betrieb Autonomie über die Nutzung und Verwendung des Humankapitals anstrebt (vgl. Hohn 1989, S. 83).

In dieser Argumentation wird zwar einerseits die Bedeutung externer, nicht unmittelbar durch den Betrieb kontrollierbarer Markt- und Konkurrenzbedingungen berücksichtigt, andererseits werden die betrieblichen Aktionen und Reaktionen nicht eindeutig durch die jeweiligen Rahmenbedingungen festgelegt. Betriebliche Strategien und Maßnahmen sind dabei als Vermittlungsprozesse zwischen den externen Rahmenbedingungen und den betriebsinternen Strukturen zu verstehen (vgl. Frick 1992, S. 106f.). Ausgehend von dieser Sichtweise kann möglicherweise erklärt werden, warum eine betriebliche Selbstbindung gegenüber den Beschäftigten, die sich notwendigerweise aus einer Arbeitskräftestrategie der Internalisierung ergibt, auch über die Zeiten einer Hochkonjunktur hinaus, also sogar in Rezessionsphasen, weiterverfolgt wird. Hartmann (1979, S. 153) vertritt die These, daß einerseits die Motivation der Beschäftigten und andererseits die Sicherung der (partiellen) Autonomie gegenüber der Umwelt durch entsprechende organisatorische Arrangements erhalten bleibt.

## 2.2 Ziele der betrieblichen Personal- und Beschäftigungspolitik

In der neueren personalwirtschaftlichen Literatur werden drei teils komplementäre, teils widersprüchliche Ziele der betrieblichen Personal- und Beschäftigungspolitik unterschieden: Die Minimierung der Arbeitskosten, eine möglichst große Flexibilität beim Einsatz von Arbeitskräften sowie die Stabilisierung bestimmter Beschäftigungsverhältnisse (vgl. Frick 1992, S. 110). Kostenminimierung meint hier nicht allein die kurzfristige Anpassung der reinen Lohnkosten an die gegebenen Produktions- und Absatzbedingungen, sondern auch die langfristige Minimierung aller mit der Rekrutierung, Qualifizierung, Beschäftigung und Entlassung in Verbindung stehenden Arbeits- und Organisationskosten. Je nach Pla-

nungshorizont und/oder personalpolitischer Risikoneigung werden Unternehmen eine kurzfristige externe Anpassungsflexibilität gegenüber einer langfristigen Minimierung der Arbeits- und Organisationskosten präferieren. Die Länge des Planungshorizonts und die personalpolitische Risikoneigung sind im wesentlichen von betrieblichen Faktoren wie etwa Größe, Konjunkturreakibilität und Kapitalintensität abhängig (vgl. Lutz 1987, S. 85-94). Eine langfristige Kostenminimierung setzt eine qualifizierte, loyale und flexibel einsetzbare Belegschaft mit einer hohen Betriebsbindung voraus (vgl. Frick 1992, S. 110). Stabile Arbeitsbeziehungen wirken in aller Regel produktivitätssteigernd und damit kosten-senkend, es besteht deshalb ein Anreiz zur Etablierung eines internen Arbeitsmarktes. Da aber gleichzeitig die Notwendigkeit zu Anpassungen an sich verändernde Rahmenbedingungen besteht, streben Unternehmen möglicherweise eine hohe externe Anpassungsflexibilität an (vgl. Buttler 1987, S. 11). "Interne Arbeitsmärkte fördern zwar die Stabilität von Beschäftigungsverhältnissen, konfliktieren jedoch partiell mit dem Erfordernis der flexiblen Anpassung bei wechselnden Marktlagen und technischen Veränderungen" (Frick 1992, S. 111).

Interne Arbeitsmärkte, die eine weitgehende Abgrenzung vom externen Arbeitsmarkt zur Folge haben, werden deshalb überwiegend in Unternehmen oder Betrieben vorzufinden sein, die aufgrund ihrer Größe, ihrer Stellung auf den relevanten Absatz- und Beschaffungsmärkten sowie der Organisation des Produktionsprozesses dazu in der Lage sind bzw. dazu gezwungen sind. Frick (1992, S. 111f.) vermutet eine derartige Internalisierung demnach in Betrieben mit geringem Anpassungserfordernis und/oder geringen Anpassungsmöglichkeiten, weil eine solche Internalisierung das betriebliche Anpassungsvermögen zumindest partiell einschränkt. Geringe Anpassungserfordernisse sieht er für Betriebe mit einer weitgehend konjunkturunabhängigen Produktion oder für Betriebe mit einer starken Stellung auf den relevanten Absatzmärkten; geringe Anpassungsmöglichkeiten dürften dagegen in kapitalintensiven Betrieben vorzufinden sein, "deren Rentabilität stark von der Funktionstüchtigkeit komplexer Produktionsanlagen abhängt, was eine stabile, eingearbeitete und loyale Belegschaft erfordert" (Frick 1992, S. 112).

Eine Implementierung von aufeinander abgestimmten Qualifizierungs- und Allokationsstrategien beeinträchtigt die externe betriebliche Anpassungsflexibilität in dem Sinne, daß die betrieblichen Möglichkeiten des externen Arbeitskräfteaus-tauschs eingeschränkt werden. Die interne Flexibilität, also die Möglichkeit innerbetrieblicher Arbeitsplatzwechsel, wird jedoch durch die Implementierung interner Arbeitsmärkte erhöht. Werden beispielsweise die Arbeitnehmer durch entsprechende betriebsspezifische Qualifizierungsmaßnahmen während ihrer Zugehörigkeit zum Betrieb auch auf anderen Arbeitsplätzen einsetzbar, dann hat dies zwar eine Einschränkung der externen Flexibilität, aber gleichzeitig eine Erhöhung der internen Flexibilität zur Folge (vgl. Frick 1992, S. 113).

Vor dem Hintergrund einer unsicheren Umwelt, die möglicherweise kurzfristige Anpassungen der Belegschaftsstärke an geänderte Rahmenbedingungen auf den Faktor- und Absatzmärkten notwendig macht und einer betrieblichen Personalpolitik, die zu einer Verstetigung der Beschäftigungsverhältnisse des Teils der Belegschaft führt, der nennenswertes betriebsspezifisches Humankapital besitzt, ist eine Spaltung der Belegschaft in eine Stamm- und eine Randbelegschaft eine mögliche Konsequenz. Die Zugehörigkeit zur Stamm- bzw. Randbelegschaft hängt von verschiedenen Faktoren ab (vgl. Mendius/Sengenberger 1976, S. 38):

- von den Knappheitsverhältnissen bestimmter Qualifikationsprofile auf dem externen Arbeitsmarkt,
- von den Entlassungs- und Rekrutierungskosten verschiedener Arbeitnehmer (-gruppen),
- von der Verhandlungsmacht verschiedener Arbeitnehmergruppen
- von dem in einzelne Arbeitnehmer(-gruppen) investierten Humankapital, wobei die Investitionsbereitschaft von der tatsächlichen oder erwarteten Betriebszugehörigkeitsdauer abhängig ist und
- von zugunsten bestimmter Arbeitnehmergruppen geltenden gesetzlichen und tarifvertraglichen Bestandsschutzregeln.

Demnach sind in der Stammebelegschaft überwiegend Arbeitnehmer mit auf dem überbetrieblichen Arbeitsmarkt stark nachgefragten Qualifikationen, Arbeitnehmer, deren betriebsspezifische Qualifizierung mit hohen Kosten verbunden war, und Arbeitnehmer, die arbeits- oder tarifvertraglich vergleichsweise stark vor Entlassung geschützt sind, zu erwarten. "In den Randbelegschaften finden sich demgegenüber jene Arbeitnehmer(-gruppen), die aufgrund ihrer permanenten Verfügbarkeit auf dem überbetrieblichen Arbeitsmarkt relativ einfach zu rekrutieren und ohne größeren Qualifizierungsaufwand einsetzbar sind, diejenigen, die relativ leicht und ohne größere finanzielle Belastungen (wie z.B. Abfindungszahlungen) entlassen werden können, sowie jene, deren Entlassung nicht bzw. nur zu unwesentlichen betrieblichen Humankapitalverlusten führt" (Frick 1992, S. 114).

### **2.3 Die Doppelfunktion interner Arbeitsmärkte**

Die Vorteilhaftigkeit aufeinander abgestimmter Qualifizierungs- und Allokationsstrategien, also interner Arbeitsmärkte, ist in zwei wesentlichen Punkten zu sehen: Zum einen ermöglichen derartige Strategien eine effiziente Personalallokation mit gleichzeitiger Schaffung und Sicherung von Humankapital. Zum ande-

ren dienen sie als Systeme der Konfliktregulierung, tragen also zur Schaffung und Sicherung von Organisationskapital bei<sup>105</sup>.

In der Humankapitaltheorie (vgl. Becker 1962, Mincer 1962) werden Ausbildung und Qualifikation der Arbeitnehmer als Investitionen, sowohl von Arbeitgebern als auch von Arbeitnehmern, interpretiert, und die Entstehung von internen Arbeitsmärkten wird folglich auf die Existenz sogenannter quasi-fixer Beschäftigungskosten zurückgeführt (vgl. Oi 1962). Der Einsatz von aufeinander abgestimmten Rekrutierungs-, Qualifizierungs- und Entlohnungspraktiken stellt also einerseits die "Versorgung" des Betriebes mit für die betrieblichen Belange notwendigen Qualifikationen sicher, und andererseits können die Investitionen in die Bildung (betriebs-)spezifischen Humankapitals aufgrund einer hinreichend langen Nutzung der Qualifikationen amortisiert werden (vgl. Frick 1992, S. 123). Eine derartige interne Qualifizierung von Arbeitskräften verursacht dann geringere Kosten als externe Beschaffung entsprechend qualifizierter Arbeitskräfte, wenn der Qualifikationserwerb als Kuppelprodukt des Arbeitsprozesses erfolgt, d.h. wenn die Qualifizierung nur geringe Verzögerungen und Beeinträchtigungen des Produktionsprozesses verursacht. Ein zusätzlicher Vorteil der internen Qualifizierung ist zudem darin zu sehen, daß im Gegensatz zu einer externen Beschaffung die Qualifizierung auf die im jeweiligen Betrieb vorhandenen Anforderungen angepaßt werden kann, "Überschußqualifikationen" also weitgehend vermieden werden können.

Neben der in den meisten Fällen kostengünstigeren Qualifikationsbereitstellung ermöglichen hierarchisch organisierte interne Arbeitsmärkte eine kostengünstige Information über die Fähigkeiten und die Leistungs-, Anpassungs- und Kooperationsbereitschaft der Beschäftigten. Es verringert sich daher das Risiko von Fehlinvestitionen in betriebliches Humankapital, und es wird eine verbesserte Nachwuchsplanung möglich.

Wenn interne Qualifizierung die Ausbildungs- und Trainingskosten reduzieren soll, dann müssen die Beschäftigten zum einen bereit sein, ihre Fertigkeiten an andere weiterzugeben, und zum anderen müssen neu rekrutierte Mitarbeiter bereit sein, die betrieblichen Normen zu akzeptieren, da dies eine entscheidende Voraussetzung für den Erwerb der am Arbeitsplatz benötigten Qualifikationen darstellt. Um die Weitergabe von Fertigkeiten an nachfolgende Mitarbeitergenerationen sicherzustellen, bedarf es der Ausschaltung des Lohnwettbewerbs. Müssen

---

<sup>105</sup> Der Begriff des Organisationskapitals stellt eine Erweiterung des Humankapitalbegriffs dar. Während ersterer den gesamten Bestand an produktiven Fähigkeiten erfaßt, die für die Höhe des tatsächlichen oder potentiellen Einkommens maßgeblich sind, beschreibt der Begriff Organisationskapital "den Bestand an (impliziten oder expliziten) Vereinbarungen und institutionellen Vorkehrungen zur langfristigen und "rücksichtsvollen" Nutzung von Humankapital" (Frick 1992, S. 117).

bereits beschäftigte Arbeitnehmer befürchten, von neu in den Betrieb eintretenden Mitarbeitern verdrängt zu werden, dann werden sie ihre Fähigkeiten möglicherweise nicht weitergeben. "In einer solchen Situation kann es durchaus im Interesse des Arbeitgebers liegen, den Lohnwettbewerb zu unterbinden und Arbeitsplatzsicherheit zu schaffen. Die Beschäftigungspraktiken eines internen Arbeitsmarktes - und hier insbesondere eine fixe Lohnstruktur, die Schaffung innerbetrieblicher Mobilitätsketten und die Beschränkung der Arbeitsplatzkonkurrenz auf die Eintrittsarbeitsplätze am unteren Ende der betrieblichen Arbeitsplatzhierarchie - sind Ausdruck dieser Bestrebungen" (Frick 1992, S. 127f., vgl. Thurow 1975, S. 81).

Empirisch findet sich sowohl auf Basis von Individual- und Unternehmensdaten als auch in verschiedenen Untersuchungsregionen (Deutschland und USA) Evidenz dafür, daß die betrieblichen Ausgaben für (arbeitsplatz-)spezifische Qualifizierungsmaßnahmen mit der Betriebsgröße steigen (vgl. Gerlach/Schasse 1988, S. 8 u. 11, Barron/Black/Loewenstein 1987, S. 82ff). Frick (1992, S. 129f.) kommt aufgrund seiner Analysen der ersten Welle des Sozio-ökonomischen Panels von 1984 zu dem Ergebnis, daß Großbetriebe signifikant mehr in die Schaffung spezifischen Humankapitals investieren als Klein- und Mittelbetriebe. Rund 25% der in Klein- und Mittelbetrieben beschäftigten Arbeitnehmer geben an, daß zur Ausübung ihrer derzeitigen Tätigkeit keine besondere Qualifikation bzw. nur eine kurze Einweisung erforderlich ist. In Großbetrieben beträgt der entsprechende Anteil dagegen nur knapp 16%. "Umgekehrt ist der Anteil der Arbeitsplätze, die eine längere Einarbeitung bzw. den Besuch besonderer Fortbildungskurse erfordern, in Großbetrieben mit 17% bzw. 23% signifikant höher als in Kleinbetrieben (13% bzw. 7%)" (Frick 1992, S. 130).<sup>106</sup>

Die Unmöglichkeit, einen Arbeitsvertrag vollständig zu spezifizieren, läßt sowohl dem Arbeitgeber als auch dem Arbeitnehmer mehrere Handlungsoptionen offen<sup>107</sup>. Während durch den Abschluß des Beschäftigungsvertrages eine Dimension des Allokationsproblems, nämlich die Abstimmung von Angebots- und Nachfragequantitäten, gelöst ist, ergibt sich aus der fehlenden Festlegung des Anstrengungsniveaus und sich ständig ändernder Umweltbedingungen für beide Parteien die Möglichkeit zu opportunistischem Verhalten. Eine Ursache für dieses Verhalten ist darin zu sehen, daß einseitiges opportunistisches Verhalten zu höherem

---

<sup>106</sup> Für das allgemeine Humankapital lassen sich dagegen nur geringfügige Unterschiede zwischen Großbetrieben einerseits und Klein- und Mittelbetrieben andererseits nachweisen. Nur der Akademikeranteil ist in Großbetrieben wesentlich höher als in Klein- und Mittelbetrieben. Im Gegensatz dazu ist der Anteil der Beschäftigten mit abgeschlossener Berufsausbildung in Klein- und Mittelbetrieben um 13 Prozentpunkte höher als in Großbetrieben.

<sup>107</sup> Zu den Besonderheiten des Arbeitsvertrages vgl. z. B.: Schmid 1987, S. 37; Duda 1987, S. 109; Becker 1985, S. 69; Dragendorf/Hering 1987, S. 134ff; Sengenberger 1987, S. 79.

Nutzen führt, wenn sich die andere Vertragspartei kooperativ verhält.<sup>108</sup> Dies gilt natürlich für beide Vertragsparteien, so daß bei vollkommener Unsicherheit über das Verhalten der Gegenseite beiderseitiges opportunistisches Verhalten die Folge ist. Situationen dieser Art, bei denen die Vertragsparteien einander mißtrauen, lassen sich als Gefangenendilemma-Situation auffassen (vgl. Sadowski 1989, S.225).

Betrachtet man das Beschäftigungsverhältnis als sich permanent wiederholendes Gefangenendilemma-Spiel, dann kommen für beide Seiten vorteilhafte Lösungen, d.h. Lösungen die zu einer entsprechenden Kooperationsrente führen, möglicherweise auch dann nicht zustande, wenn dies von beiden Seiten "eigentlich" gewollt ist. Auch wenn zu einem bestimmten Zeitpunkt eine kooperative Situation erreicht ist, die eine Kooperationsrente sicherstellt, ist nicht gewährleistet, daß diese Lösung dauerhaft bestehen bleibt. Axelrod (1984, S. 126ff) leitet aus einer Analyse sich wiederholender Gefangenendilemma-Spiele vier grundsätzliche Möglichkeiten zur Herstellung dauerhafter Kooperation ab:

#### a) Erhöhung der Bedeutung der Zukunft

Die Wahrscheinlichkeit einer stabilen wechselseitigen Kooperation wächst mit der Bedeutung der Zukunft in Relation zur Gegenwart. Die Bedeutung zukünftiger Entwicklungen wird um so wichtiger, je dauerhafter die Interaktion der beteiligten Parteien ist. Dies ist darauf zurückzuführen, daß der Anreiz zu opportunistischem oder nicht-kooperativem Verhalten abnimmt, wenn eine Partei damit rechnen muß, zukünftig mit einem nicht-kooperativen Verhalten der Gegenpartei konfrontiert zu werden. Besteht etwa für beide Vertragsparteien (Arbeitgeber und Arbeitnehmer) aufgrund von nicht transferierbarem spezifischen Humankapital (Arbeitnehmer) und hohen Investitionen in betriebsspezifisches Humankapital (Arbeitgeber) ein Interesse an einem langfristigen Beschäftigungsverhältnis, dann wird die Verwirklichung einer kooperativen Verhaltensoption wahrscheinlicher.

#### b) Veränderung der Auszahlungen

Eine Verringerung der Auszahlungen bzw. eine Erhöhung der Kosten bei Abweichen von der kooperativen Verhaltensweise führt zu einer geringeren Wahrscheinlichkeit von unkooperativen Verhaltensweisen. In der Bundesrepublik belegen beispielsweise das Arbeitsrecht und Tarifverträge bestimmte nicht-kooperative Verhaltensweisen mit hohen Kosten. Wegen der Besonderheiten des Arbeitsvertrages ist es nicht möglich, durch Regelungen auf überbetrieblicher Ebene alle denkbaren Eventualitäten von nicht kooperativem Verhalten mit entsprechend hohen Kosten zu belegen. Ergänzend zu justiziablen Bestandteilen von

---

<sup>108</sup> Kooperatives Verhalten meint hier, daß sich eine Vertragspartei an möglicherweise getroffene Absprachen oder implizite Vereinbarungen hält.

Verträgen müssen Regeln implementiert werden, die nicht kooperatives Verhalten verstärkt sanktionieren bzw. kooperatives Verhalten belohnen.<sup>109</sup>

#### c) Erhöhung der Verantwortlichkeit

In einer Beziehung, in der die beteiligten Parteien sich um das Wohlergehen der anderen kümmern, ist Kooperation eher realisierbar. Unabhängig davon, ob Altruismus oder Kostenüberlegungen als Motiv unterstellt werden, kann von verantwortungsvollem Verhalten einer Partei ein Signal ausgehen, das sich förderlich auf das kooperative Verhalten der anderen Partei auswirkt. So ist zu vermuten, "daß jüngere und neu eintretende Arbeitnehmer den Umgang mit Älteren und (Schwer-)Behinderten als Indiz dafür werten werden, was sie selbst erwarten können. Das beschränkt Arbeitgeber in ihren Verhaltensweisen auf solche, die auch von den Jüngeren und Abwanderungsfähigen als "gerecht" und "fair" empfunden werden" (Frick 1991, S. 311). Da diese Verhaltensweisen aber nur den Charakter impliziter Vereinbarungen haben, "kann die (Weiter-) Beschäftigung schwerbehinderter Belegschaftsmitglieder unabhängig von gesetzlichen Auflagen als "freiwilliger moralökonomischer Akt" von Unternehmen verstanden werden, da Reziprozitätsnormen<sup>110</sup> weitgehend unabhängig von gesetzlichen Regelungen die betrieblichen Verfügungs- und Nutzungsrechte einschränken" (ebenda, S. 311).

Ähnliche Effekte sind möglicherweise durch die Weiterbeschäftigung und Therapie von Suchtkranken zu erwarten.

#### d) Verbesserung des Erinnerungsvermögens

Beschränkte Rationalität, begrenzte Informationskapazität und falsche Wahrnehmung von Verhaltensweisen können ein Erkennen oder Wiedererkennen der Identität oder der Handlungen der anderen Partei verhindern. Dieses Problem tritt dann verstärkt auf, wenn die Folgen einer Handlungsweise nicht offensichtlich

---

<sup>109</sup> Gesetze, die nicht-kooperatives Verhalten verhindern sollen, verfehlen möglicherweise ihre Wirkung: Sadowski (1989, S. 225) verweist darauf, daß selbst arbeitnehmerseitig gewonnene Kündigungsschutzprozesse nur selten zur Wiedereinstellung führen und beispielsweise Dienst nach Vorschrift bereits als Arbeitsverweigerung gilt. Gottfredson/White (1981, S. 497f.) zeigen Bedingungen auf, unter denen implizite Übereinkünfte effizienter sein können als formal festgelegte.

<sup>110</sup> Kohli/Rosenow/Wolf (1983, S. 29) definieren die Reziprozitätsnormen als "... the basic concept of justice and equity under which individuals organize their economic action. ... On the one hand, this involves the expectation the utilization of labor in the firm does not endanger the lifetime protection of the capacity for work. On the other hand it is felt that the employees furnish the firm with an investment, based on their continuous performance, their willingness to accept responsibility, their reliability, etc. - i.e. especially the non-contractual elements of work - for which the firm will reward them with special benefits if their performance capacity should diminish some day."

sind oder die Konsequenzen weit in der Zukunft liegen und damit eine Kontrollierbarkeit erschwert ist. So können beispielsweise betriebliche Interessenvertretungen aufgrund in der Vergangenheit gemachter Erfahrungen dafür sorgen, arbeitgeberseitige Verstöße gegen implizite Vereinbarungen sichtbar zu machen, während diese jüngeren Mitarbeitern ansonsten möglicherweise verborgen bleiben; auf diese Weise erhöhen sich die Kosten der Nichteinhaltung impliziter Vereinbarungen. Andererseits können betriebliche Interessenvertretungen durch den Arbeitgeber dazu genutzt werden, bestimmte "unpopuläre" Entscheidungen ohne nennenswerten Reputationsverlust durchzusetzen.

Die Implementation von aufeinander abgestimmten Einstellungs-, Ausbildungs- und Entlohnungspraktiken enthält im wesentlichen Elemente dieser vier allgemeinen Grundsätze zur Herstellung von Kooperation, d.h. interne Arbeitsmärkte ermöglichen die Bildung von "Vertrauens- oder Organisationskapital". Wenn der rücksichtsvolle Umgang mit Älteren oder Schwerbehinderten von jüngeren oder neu in den Betrieb eintretenden Arbeitnehmern als Indiz dafür gewertet wird, was sie selbst erwarten können, dann führt dies bei den jüngeren und neuen Arbeitnehmern möglicherweise zu kooperativem Verhalten, woraus Produktivitätssteigerungen resultieren.

#### 2.4 Zusammenfassung der Hypothesen<sup>111</sup>

Frick (1992, S. 151ff) formuliert auf Basis der oben skizzierten theoretischen Überlegungen folgende Hypothesen:

1. Erfordert die Schaffung und Sicherung von Organisationskapital eine langfristige und rücksichtsvolle Nutzung von Arbeitskraft, dann ist zu erwarten, daß die (Weiter-)Beschäftigung Schwerbehinderter in geringerem Maße von gesetzlichen (Bestandsschutz-)Regelungen als vielmehr von freiwilligen organisationskostenreduzierenden Beschäftigungsstrategien der Unternehmen abhängt. Frick folgert hieraus,
  - 1.1 daß Schwerbehinderte weder hinsichtlich ihres Einkommens noch hinsichtlich ihres beruflichen Status bzw. ihrer Tätigkeit diskriminiert werden,
  - 1.2 daß Schwerbehinderte eine signifikant längere Betriebszugehörigkeitsdauer aufweisen als Nichtbehinderte und
  - 1.3 daß "intern rekrutierte" Schwerbehinderte bei Kontrolle von Alter, Geschlecht und Qualifikation eine signifikant größere Betriebszugehörigkeitsdauer haben als "extern rekrutierte" Schwerbehinderte.<sup>112</sup>

---

<sup>111</sup> Die Hypothesenblöcke 1-4 werden von Frick (1992, S. 151ff) übernommen. Aufgrund der synthetischen Stichprobe werden jedoch auch Hypothesen, die den Zusammenhang von Schwerbehindertenquote und Produktivität beinhalten, testbar (Hypothese 5).

2. Entspricht die (Weiter-)Beschäftigung gesundheitlich beeinträchtigter und (schwer-)behinderter Arbeitnehmer den Reziprozitätserwartungen der Belegschaft, und werden arbeitgeberseitige Verstöße gegen diese Erwartungen von den betrieblichen Interessenvertretungen sanktioniert, dann kann erwartet werden,
  - 2.1 "daß Betriebe oder Dienststellen mit Betriebs- bzw. Personalrat und/oder Schwerbehindertenvertretung eine signifikant höhere Istquote aufweisen als Betriebe und Dienststellen ohne gewählte Interessenvertretung(en) und
  - 2.2 daß eine "flächendeckende" Verbreitung dieser Interessenvertretungen einen nennenswerten Anstieg der Istquote zur Folge hätte" (Frick 1992, S. 151).
3. In einer hohen durchschnittlichen Betriebszugehörigkeitsdauer und einer niedrigen Fluktuationsrate sind mögliche Indikatoren für eine rücksichtsvolle und langfristige Nutzung von Arbeitskraft bzw. einen hohen Bestand an Organisationskapital zu sehen. Hieraus kann
  - 3.1 auf einen positiven Zusammenhang zwischen Schwerbehindertenquote und Betriebszugehörigkeitsdauer und
  - 3.2 einen negativen Zusammenhang zwischen Fluktuationsrate und Schwerbehindertenanteil geschlossen werden.
  - 3.3 Die Stabilisierung von Beschäftigungsverhältnissen läßt sich unter anderem durch die Zahlung von "Senioritätslöhnen" erreichen, so daß zu vermuten ist, daß mit steigendem durchschnittlichen Pro-Kopf-Einkommen auch die Schwerbehindertenquote zunimmt.
4. Ob und inwieweit Betriebe oder Unternehmen dazu in der Lage sind bzw. die Bereitschaft besitzen, langfristige Beschäftigungsbeziehungen zu etablieren, hängt von mehreren betriebsinternen und -externen Faktoren ab. Frick spezifiziert in diesem Zusammenhang c.p. folgende Hypothesen:
  - 4.1 Mit zunehmender Betriebsgröße steigen die "monitoring costs", und die Notwendigkeit der internen Stabilisierung nimmt zu. Es ist folglich zu erwarten, daß mit zunehmender Größe die Betriebszugehörigkeitsdauer steigt und die Fluktuationsrate abnimmt.
  - 4.2 Eine starke wirtschaftliche Konzentration ermöglicht den Betrieben eine Politik der internen Stabilisierung, so daß auf einen positiven Zusammenhang zwischen Betriebszugehörigkeitsdauer und Konzentration bzw. auf einen ne-

---

112 "Intern rekrutiert" meint in diesem Zusammenhang, daß die Behinderung während der Beschäftigung an der derzeitigen Arbeitsstelle eingetreten ist; "extern rekrutiert" bedeutet, daß Schwerbehinderte bereits als "schwerbehindert" eingestellt worden sind.

gativen Zusammenhang zwischen Fluktuationsrate und Konzentration geschlossen werden kann.<sup>113</sup>

- 4.3 Weil ein Anstieg der Arbeitslosigkeit eine Reduktion der Fluktuationskosten, d.h. der Alternativkosten der Vertragserneuerung bewirkt, sinkt die Betriebszugehörigkeitsdauer und steigt die Fluktuationsrate mit der Arbeitslosenquote.
- 4.4 In wachsenden Betrieben und Branchen ist die Zahl sogenannter "Matching-Prozesse" größer als in schrumpfenden Betrieben und Branchen, so daß in wachsenden Betrieben und Branchen eine höhere Fluktuationsrate als in schrumpfenden Betrieben und Branchen erwartet werden kann.
- 4.5 Die höhere Elastizität der betrieblichen Nachfrage nach un- bzw. geringqualifizierten Arbeitskräften hat eine geringere Betriebszugehörigkeitsdauer und eine höhere Fluktuationsrate dieser Arbeitskräfte zur Folge.
- 4.6 Die Verfügbarkeit gesellschaftlich legitimierter bzw. akzeptierter Alternativrollen läßt für Frauen eine geringere durchschnittliche Betriebszugehörigkeitsdauer und höhere Fluktuationsrate als bei den Männern erwarten.
5. Wenn eine rücksichtsvolle und langfristige Nutzung von Arbeitskraft zum Entstehen einer Kooperationsrente führen kann, und eine hohe Schwerbehindertenquote Ausdruck einer derartigen Personalpolitik ist, dann kann in Betrieben mit einem hohen Schwerbehindertenanteil c.p. mit einer höheren Arbeitsproduktivität als in Betrieben mit niedrigem oder keinem Schwerbehindertenanteil gerechnet werden.

### 3. Ausgewählte empirische Ergebnisse von Frick<sup>114</sup>

#### 3.1 Methodisches Vorgehen

Angesichts des Datenmangels über Betriebe bzw. Unternehmen<sup>115</sup> greift Frick bei den empirischen Tests seiner Hypothesen überwiegend auf Individual- und Branchendaten zurück. Lediglich die Analyse des Einflusses von betrieblichen Interessenvertretungen auf die Schwerbehindertenquote erfolgt auf Basis von Betriebsdaten. Obwohl Frick idealerweise zum adäquaten Testen seiner Hypothe-

---

<sup>113</sup> Eine starke Konzentration in einer Region und/oder Wirtschaftszweig kann jedoch auch zu einer monopsonistischen Situation auf dem relevanten Arbeitsmarkt führen. In einem solchen Fall könnte der Zusammenhang zwischen Fluktuationsrate und Konzentration, bzw. Betriebszugehörigkeitsdauer und Konzentration nivelliert oder sogar umgekehrt sein, da der Bruch impliziter Vereinbarungen durch den Arbeitgeber zwar einen Reputationsverlust zur Folge hätte, aber aufgrund fehlender Alternativen der Arbeitsanbieter kaum negative Sanktionen befürchtet werden müßten.

<sup>114</sup> vgl. hierzu Frick 1992, S. 155ff.

<sup>115</sup> Für eine Beschreibung der Datensituation siehe Frick 1992, S. 157-160.

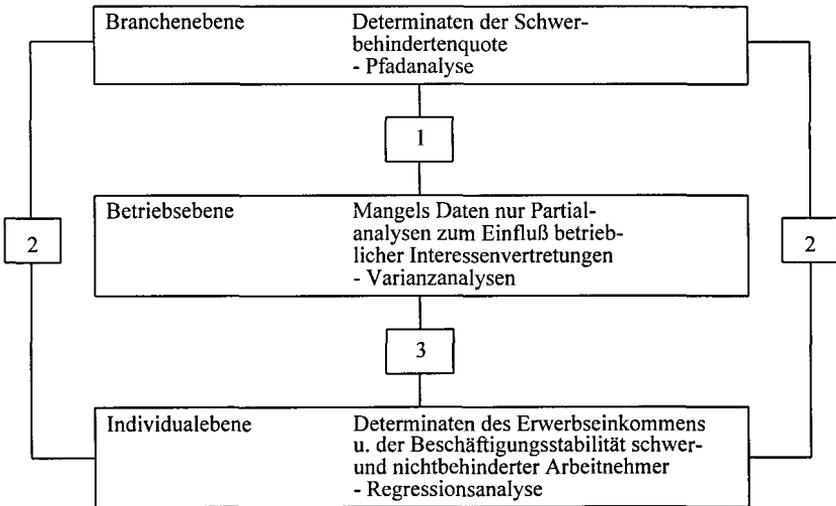
sen Unternehmensdaten benötigen würde, versucht er eine Aussage über Betriebs- und/oder Unternehmensverhalten durch eine simultane Analyse von Individual- und Branchendaten zu treffen. "Unter der Annahme, daß Zusammenhänge, die sich auf der Branchen- und Individualebene nachweisen lassen, beim Vorliegen entsprechender Daten auch für die Betriebsebene dokumentiert werden können, werden ... veröffentlichte, aber auch in aller Regel unzugängliche Daten sowohl aus der amtlichen Statistik als auch aus der Umfrageforschung genutzt, um die ... skizzierte Fragestellung empirisch beantworten zu können. Durch die simultane Analyse von Branchen- und Individualdaten bei gleichzeitiger Verwendung der wenigen verfügbaren Betriebsdaten ist es möglich, das Risiko eines sogenannten "ökologischen Fehlschlusses" zu minimieren ... Darunter versteht man den irrtümlichen Schluß von Eigenschaften, die beispielsweise bei Wirtschaftszweigen beobachtet werden, auf die entsprechenden Eigenschaften der Betriebe innerhalb dieser Branchen ..." (Frick 1992, S. 170).

Diese Vorgehensweise wird als "Quasi-Mehrebenenanalyse"<sup>116</sup> bezeichnet. Übersicht 2 zeigt eine schematische Darstellung dieser Vorgehensweise:

---

<sup>116</sup> Vgl. Fuchs et. al. 1978, S. 421; Huinink 1989.

## Übersicht 2: Vorgehensweise bei der empirischen Analyse von Frick



- 1 Empirische Analyse von Branchendaten zur Erklärung betrieblichen Verhaltens.
- 2 Plausibilitäts- bzw. Kompatibilitätsüberprüfungen.
- 3 Empirische Analyse von Individualdaten zur Erklärung betrieblichen Verhaltens.

Quelle: Frick 1992, S. 171

Die komplementäre Analyse von Branchen- und Individualdaten unterschiedlicher Quellen ersetzt zwar nicht die theoretisch erforderliche Analyse von Betriebsdaten, es kann jedoch eine vorläufige Substitution erreicht werden (vgl. Frick 1992, S. 172). Der Versuch mittels einer synthetischen Stichprobe, einerseits die von Frick ermittelten Ergebnisse mit Daten auf Unternehmensebene zu reproduzieren und andererseits weitergehende Analysen durchzuführen, ermöglicht eine erste Abschätzung, inwieweit die Methode der "Quasi-Mehrebenenanalyse" sich zur Analyse von Unternehmens- und/oder Betriebsverhalten eignet.

### 3.2 Darstellung einzelner Ergebnisse<sup>117</sup>

In der empirischen Analyse der "Determinanten der (Nicht-)Beschäftigung Schwerbehinderter" geht Frick in drei Schritten vor: Zunächst geht er der Frage nach der Existenz einer Diskriminierung Schwerbehinderter bezüglich Einkommen und Beschäftigung nach. In einem zweiten Schritt untersucht er den Einfluß der betrieblichen Interessenvertretung auf die Schwerbehindertenbeschäftigung, bevor dann abschließend die Bestimmungsgründe der betrieblichen Schwerbehindertenquote analysiert werden.

Zur Beantwortung der Frage, inwieweit Schwerbehinderte bezüglich ihres Einkommens diskriminiert werden, greift Frick auf den in der Literatur oft nachgewiesenen Zusammenhang zwischen Betriebszugehörigkeitsdauer und Einkommen zurück (vgl. z.B.: Blossfeld/Hannan/Schömann 1988, 1989; Brüderl 1989). Auf Basis der Daten der ersten Welle des Sozio-ökonomischen Panels<sup>118</sup> von 1984 zeigt Frick, daß schwerbehinderte Arbeitnehmer ein um 8% höheres monatliches Bruttoeinkommen beziehen als Nichtbehinderte. Er führt diesen Befund auf die unterschiedliche Altersstruktur von Schwer- und Nichtbehinderten zurück, weshalb dieser Befund mit der Wirkungsweise altersabhängiger Entlohnungsformen erklärt werden kann.<sup>119</sup>

Zur Erklärung der empirisch beobachteten Varianz der individuellen Erwerbseinkommen mittels ökonomischer Verfahren wird in der mikroökonomischen Arbeitsmarktforschung häufig auf aus der Humankapitaltheorie abgeleitete "Einkommensfunktionen" zurückgegriffen, die sich in vielen empirischen Untersuchungen gut bewährt haben (vgl. Psacharopoulos 1985, S. 583ff). In der klassischen Variante dieser Funktion (vgl. Mincer 1974) wird zur Erklärung nur auf die Schulbildung und die Berufserfahrung zurückgegriffen (vgl. Blien 1986, S. 28). In neueren Arbeiten werden darüber hinaus auch die Betriebszugehörigkeitsdauer und die Phasen der Erwerbsunterbrechung mit einbezogen.

Die Regressionskoeffizienten von Schulbildung, Berufserfahrung und Betriebszugehörigkeitsdauer können als "Rentabilitäts- bzw. Ertragsraten" interpretiert werden, "d.h. als prozentuale Zunahme des Arbeitseinkommens nach einer einjährigen Zunahme der (Aus-)Bildungsdauer, der Berufserfahrung oder der Betriebszugehörigkeit. Demgegenüber ist der Koeffizient der letztgenannten Variablen als "Abschreibungsrate", d.h. als prozentuale Veränderung des erzielbaren Arbeitseinkommens nach einjähriger Unterbrechung der Erwerbstätigkeit, inter-

<sup>117</sup> Vgl. Frick 1992, S. 172ff.

<sup>118</sup> Zur Konzeption vgl. z.B.: Hanefeld, U. 1984, 1985.

<sup>119</sup> Frick verweist jedoch auch auf eine weitere mögliche Erklärung für diesen Befund: Während der Frauenanteil unter den Arbeitnehmern insgesamt bei knapp 40% liegt, beträgt er bei den Schwerbehinderten nur rund 27%.

pretierbar" (Frick 1992, S. 177). Frick schätzt auf Basis der ersten Welle des Sozio-ökonomischen Panels eine Einkommensfunktion.

$$\ln Y = \beta_0 + \beta_1 \text{BIL} + \beta_2 \text{EXP} + \beta_3 \text{EXP}^2 + \beta_4 \text{BZD} + \beta_5 \text{BZD}^2 \\ + \beta_6 \text{NET} + \beta_7 \text{SEX} + \beta_8 \text{SB} + u$$

mit<sup>120</sup>

lnY: logarithmierter Bruttostundenlohn bzw. -gehalt

BIL: Schul- und Berufsausbildung

EXP: Berufserfahrung

EXP<sup>2</sup>: quadrierte Berufserfahrung

BZD: Betriebszugehörigkeitsdauer

BZD<sup>2</sup>: quadrierte Betriebszugehörigkeitsdauer

NET: Phasen der Nichterwerbstätigkeit

SEX: Geschlecht

SB: Gesundheitszustand.

Die Schätzung der Regressionsfunktion deutet nicht auf eine Einkommensdiskriminierung Schwerbehinderter hin. Vielmehr ist der Koeffizient der Variablen Gesundheitszustand insignifikant, während die anderen Variablen in der Regressionsfunktion das theoretisch erwartete Vorzeichen aufweisen und mit empirischen Ergebnissen aus anderen Studien übereinstimmen.

In einem komplexen Pfadmodell geht der Autor der Frage nach, inwieweit die betriebliche Schwerbehindertenquote von den theoretisch hergeleiteten Faktoren beeinflusst wird. In einer Regressions-schätzung, in der die Faktoren Betriebsgröße, Beschäftigungsstabilität, Arbeitslosenquote, Pro-Kopf-Einkommen, Fluktuationsrate, (allgemeines) Humankapital, Frauenquote, Beschäftigungsentwicklung und Konzentration als unabhängige Variablen berücksichtigt werden, weisen lediglich die Variablen Beschäftigungsstabilität (positiv) und Fluktuationsrate

---

<sup>120</sup> Die erklärenden Variablen werden folgendermaßen operationalisiert: Die Schul- und Berufsausbildung (BIL) ist die Summe der Jahre, die ein Individuum zum Erreichen des jeweiligen Schulabschlusses und der Berufsausbildung aufwenden mußte. Die Berufserfahrung (EXP) wird gemessen als Summe der Jahre, die ein Individuum ab dem sechzehnten Lebensjahr voll- oder teilzeitbeschäftigt war. Zur Operationalisierung der Betriebszugehörigkeitsdauer (BZD) wird die Beschäftigungsdauer beim derzeitigen Arbeitgeber verwendet. Die Phasen der Erwerbsunterbrechung (NET) werden als Summe der Jahre, in denen ein Individuum entweder arbeitslos, in Rente oder im eigenen Haushalt tätig war, gemessen. Die Variable Gesundheitszustand (SB) geht als Dummy-Variable mit den Ausprägungen 0 (nicht behindert) und 1 (schwerbehindert) in die Regressionsgleichung ein.

(negativ) einen signifikanten Einfluß auf die Schwerbehindertenquote auf, wobei die Varianzaufklärung knapp 77% beträgt. Die anderen Variablen haben keinen direkten, signifikanten Einfluß auf die Schwerbehindertenquote, determinieren aber die Beschäftigungsstabilität und die Fluktuationsrate. Die Beschäftigungsstabilität wird durch die Betriebsgröße positiv und durch die Beschäftigungsentwicklung, das Humankapital, die Arbeitslosenquote sowie die Frauenquote negativ beeinflusst. Für die Fluktuationsrate zeigt Frick einen positiven Einfluß der Arbeitslosenquote und der Beschäftigungsentwicklung sowie einen negativen Einfluß von Betriebsgröße, Humankapital, Frauenquote und Konzentration.

Frick interpretiert diese Ergebnisse dahingehend, daß er die Beschäftigung Schwerbehinderter als ein Kuppelprodukt einer Internalisierungsstrategie auffaßt. Aufgrund der gefundenen Ergebnisse konnten die aufgestellten Hypothesen auf Basis von Individual- und Branchendaten weitgehend bestätigt werden. Mit dem zur Verfügung stehenden Datenmaterial war es jedoch nicht möglich, den direkten Einfluß betrieblicher Internalisierungsstrategien und damit den Einfluß der (Weiter-)Beschäftigung von Schwerbehinderten auf die Produktivität empirisch zu testen. Wenn die (Weiter-)Beschäftigung Schwerbehinderter ein Kuppelprodukt der Internalisierungsstrategie ist, dann müßten durch Verfolgung einer derartigen Strategie Effizienzvorteile beobachtet werden können. Ungewöhnlich hoch scheint bei den kurz skizzierten Ergebnissen zu den Determinanten der betrieblichen Schwerbehinderten die errechnete Varianzaufklärung im Regressionsmodell. Zum einen arbeitet Frick mit Branchendaten, und zum anderen werden die Branchendaten aus fünf aufeinander folgenden Jahren in einem gepoolten Sample zusammengefaßt. Auf Branchenebene dürfte die Veränderung der unabhängigen und abhängigen Variablen von einem Jahr zum anderen relativ gering ausfallen, da durch die Aggregation gegenläufige Effekte nivelliert werden. Für die im Regressionsansatz verwendeten Anteilswerte sind mit hoher Wahrscheinlichkeit keine Quoten von Null vorhanden (Schwerbehinderten- und Frauenquote), die aber auf Unternehmensebene sicherlich zu beobachten sind. Die Interpretation von Branchendaten als repräsentatives Unternehmen erscheint daher als zweifelhaft.

Die unten dargestellten Regressionsergebnisse auf Basis der synthetischen Stichprobe beinhalten einen direkten Schätzansatz zwischen Bruttowertschöpfung pro Beschäftigtem und der Schwerbehindertenquote pro Unternehmen sowie anderer aus der Theorie Interner Arbeitsmärkte ableitbarer Größen, so daß auch deren Einfluß auf die Bruttowertschöpfung kontrolliert werden kann.

## 4. Konstruktion der synthetischen Unternehmensstichprobe

### 4.1 Auswahl der Variablen und Aggregatinformationen

Zum Testen der in Abschnitt 2.4 dargestellten Hypothesen wäre es wünschenswert, wenn die dort verwendeten Größen unmittelbar gemessen und damit in den entsprechenden ökonomischen Modellen verwendet werden könnten. Wie die Ausführungen von Frick (1992) jedoch zeigen, sind selbst auf einem hohen Aggregationsniveau "große" Operationalisierungsanstrengungen notwendig, um überhaupt empirische Tests durchführen zu können. Da bei der Konstruktion einer synthetischen Stichprobe ebenfalls auf Aggregatinformationen zurückgegriffen wird, besteht auch hier ein Operationalisierungsbedarf.

Der einfachste Weg zur Rekonstruktion und "Überprüfung" der Ergebnisse von Frick mit Daten auf Unternehmensebene ist die Verwendung bzw. Disaggregation der dort verwendeten Aggregatdaten. Gegen die Wahl dieses Weges sprechen jedoch im wesentlichen drei Gründe:

- (1) In der Arbeit von Frick werden aggregierte Daten aus der Beschäftigtenstatistik verwendet, die aus dem Zeitraum von 1982-1986 stammen. Diese Informationen sind mit allen Problemen behaftet, die in Zusammenhang mit der Beschäftigtenstatistik diskutiert werden. Hierzu zählen insbesondere die Betriebsnummernvergabe und die Nichterfassung von Einkommen oberhalb der Beitragsbemessungsgrenze.
- (2) Mit der Arbeitsstättenzählung von 1987, deren Ergebnisse mittlerweile veröffentlicht sind, liegt eine relativ neue Vollerhebung der bundesrepublikanischen (alte Bundesländer) Arbeitsstätten und Betriebe vor, so daß getroffene empirische Aussagen zumindest für den Zeitpunkt der Zählung möglicherweise eine größere Zuverlässigkeit besitzen.
- (3) Das zur Erzeugung der synthetischen Stichprobe eingesetzte Verfahren bietet die Möglichkeit, Informationen aus mehreren Datenquellen miteinander zu verknüpfen, so daß über die von Frick getesteten Hypothesen hinaus Zusammenhänge zwischen Schwerbehindertenquote und Produktivität testbar werden.

Die Auswahl alternativer Datenquellen als Grundlage für den Generierungsprozeß hat aber auch zur Folge, daß einige von Frick in die empirische Analyse einbezogene Variable nicht in die synthetische Stichprobe aufgenommen werden können.

#### 4.1.1 Die Auswahl der Aggregatinformationen

Ein vorrangiges Ziel bei der Bereitstellung von Daten für empirische Hypothesentests ist eine möglichst repräsentative Abbildung der relevanten Grundgesamtheit. Im Fall von Unternehmensdaten wird Repräsentativität meist durch

Vergleich mit Branchen- und Betriebsgrößenaggregaten beurteilt bzw. bei vorliegenden größeren Abweichungen durch eine entsprechende Gewichtung hergestellt.

In einem ersten Schritt ist deshalb die Auswahl von Aggregatinformationen davon abhängig, inwieweit ein möglichst repräsentatives Bild der Grundgesamtheit bezüglich Betriebsgröße und Branche sowohl für die Zahl der Betriebe als auch für die Zahl der Beschäftigten erreicht werden kann. Vor dem Hintergrund dieser Zielsetzung erscheint die Wahl der Arbeitsstättenzählung von 1987 als Ausgangspunkt des Generierungsprozesses geeignet. Ausgehend von dieser Quelle können dann weitere Quellen, die jeweils eine "gute" Erfassung anderer Größen darstellen, zur Generierung weiterer Variablen herangezogen werden. Die Verwendung unterschiedlicher Datenquellen wird jedoch durch zum Teil sehr unterschiedliche Branchen- und Betriebsgrößenklassengliederungen erschwert. Die synthetische Stichprobe baut auf den in der Arbeitsstättenzählung verwendeten Branchen- und Betriebsgrößenklassengliederungen auf. Die zusätzlich verwendeten Datenquellen werden auf diese Gliederungsebene umgerechnet.

Nutzt man die Aggregatdaten der Arbeitsstättenzählung als Hauptdatenquelle, dann kann zwischen einer Analyse bzw. Disaggregation auf Arbeitsstätten- und auf Unternehmensebene gewählt werden. Wie in Abschnitt III.2 begründet, wird als Analyseebene hier das Unternehmen gewählt. Für die Wahl dieser Analyseinheit spricht zusätzlich, daß die vom Deutschen Institut für Wirtschaftsforschung veröffentlichten Zeitreihen zur Bruttowertschöpfung und zum Bruttoanlagevermögen für das Verarbeitende Gewerbe auf Betriebsebene erhoben werden und eine Verbindung mit den Daten auf Arbeitsstättenebene weitaus größere Differenzen liefert als eine Verknüpfung mit Daten auf Unternehmensebene aus der Arbeitsstättenzählung. In den vom DIW veröffentlichten Zeitreihen werden für 1987 im Verarbeitenden Gewerbe in Größenklassen von 20 und mehr Beschäftigten 7.005.100 Erwerbstätige ausgewiesen. Im Vergleich dazu liefert die Arbeitsstättenzählung auf Unternehmensebene eine Zahl von 7.092.928 Beschäftigten und auf Arbeitsstättenebene von 6.767.430 Beschäftigten. Die Zahlen auf Unternehmensebene weichen somit weniger von den DIW-Ergebnissen ab als die Zahlen auf Arbeitsstättenebene.

Die für die Analyse der Determinanten der betrieblichen Schwerbehindertenbeschäftigung benötigten Schwerbehindertenquoten werden vom Institut für Arbeitsmarkt- und Berufsforschung zum einen gegliedert nach Branchen und zum anderen gegliedert nach Betriebsgrößenklassen veröffentlicht<sup>121</sup>. Durch das relativ hohe Aggregationsniveau der Variable "Branche" (Hauptabteilung) ist die

---

<sup>121</sup> Diese Zahlen beruhen auf der Auswertung der jährlich obligatorisch abzugebenden Arbeitgeberanzeigen für Betriebe mit 16 und mehr Beschäftigten.

zugrundeliegende Analyseebene (Unternehmen, Betrieb oder Arbeitsstätte) von untergeordneter Bedeutung.

#### 4.1.2 Die Auswahl der Variablen (Operationalisierung)

Wählt man die Arbeitsstättenzählung von 1987 als Ausgangspunkt für die Konstruktion der synthetischen Unternehmensstichprobe, dann steht die Bildung einer vor diesem Datensatz repräsentativen Stichprobe als Ausgangspunkt für weitere Simulationen im Vordergrund. Wie oben bereits dargestellt, wird in der Literatur meist Repräsentativität bezüglich Branchenzugehörigkeit und Betriebsgrößenklasse angestrebt. Dabei wird entweder die Beschäftigtenzahl oder die Zahl der Betriebe in den entsprechenden Größenklassen und Branchen angepaßt. In der hier generierten Stichprobe wird die Repräsentativität sowohl bezüglich der Zahl der Beschäftigten als auch der Zahl der Betriebe in der jeweiligen Größenklasse und Branche realisiert.

#### Wirtschaftszweig

In der Arbeitsstättenzählung wird zwischen 10 Abteilungen, 37 Unterabteilungen, 196 Gruppen, 479 Untergruppen und 651 Klassen unterschieden (vgl. Statistisches Bundesamt 1989, Fachserie 2, Heft 1, S. 26). Die Zuordnung zu den einzelnen Kategorien erfolgt dabei aufgrund des "Schwerpunkts der wirtschaftlichen Tätigkeit". Mit zunehmender Disaggregation der Variable "Wirtschaftszweig" nimmt die Breite der veröffentlichten Variablen ab, bzw. es wird nach bestimmten Merkmalen nicht mehr differenziert. Die Zahl der Unternehmen, die Zahl der Beschäftigten (darunter diejenigen mit Löhnen und Gehältern) sowie der Lohnsumme und des Durchschnittslohns pro Beschäftigtem in der jeweiligen Betriebsgrößenklasse werden nur auf der Ebene der 196 Gruppen veröffentlicht. Die Zahl der nicht veröffentlichten Werte, beispielsweise für den Durchschnittslohn, nimmt mit zunehmender Disaggregation zu. Auf Ebene der Gruppen müßten bereits viele Werte konstruiert werden.<sup>122</sup> Um dieses Problem weitgehend zu vermeiden, wird die Variable "Wirtschaftszweig" nur auf der Ebene der 9 Unterabteilungen des Verarbeitenden Gewerbes erzeugt. Da als Analyseebene das "Unternehmen" gewählt wurde, dessen "Schwerpunkt der wirtschaftlichen Tätigkeit" bei einer sehr tiefen Differenzierung der Wirtschaftszweige zu einer falschen Klassifizierung eines Großteils der Beschäftigten führen kann und die Wahrscheinlichkeit einer Fehlklassifikation bei geringerer Disaggregation sinkt, erscheint dieses Vorgehen gerechtfertigt.

---

<sup>122</sup> Zu den Geheimhaltungsvorschriften vgl. Statistisches Bundesamt, Fachserie 2, Heft 1, S. 23.

Die Variable "Wirtschaftszweig/Branche" dient in erster Linie zur Verknüpfung mit anderen Aggregatinformationen, die zur Generierung der synthetischen Stichprobe herangezogen werden. Dabei ist eine direkte Verknüpfung nicht in allen Fällen möglich, da zum Teil andere Branchengliederungen verwendet werden. Um die vom Deutschen Institut für Wirtschaftsforschung veröffentlichten Daten zum Anlagevermögen und zur Bruttowertschöpfung im Verarbeitenden Gewerbe nutzen zu können, muß die dort gewählte Gliederung (Systematik der Wirtschaftszweige, Fassung für die Statistik im Produzierenden Gewerbe) an die Gliederung in der Arbeitsstättenzählung angepaßt werden.

### **Arbeitnehmer/Beschäftigte**

Zu den Arbeitnehmern gehören Beamte, Richter, Angestellte, Facharbeiter, Gesellen, sonstige Arbeiter und Auszubildende. Es handelt sich jedoch nur um einen Teil der Beschäftigten. Zu den Beschäftigten eines Unternehmens zählen zudem noch tätige Inhaber und unbezahlt mithelfende Familienangehörige (vgl. Statistisches Bundesamt 1991, Fachserie 2, Heft 10, S. 7). Zusätzlich zu den Arbeitnehmern insgesamt wird auch die Zahl der Arbeitnehmer mit Löhnen und Gehältern ausgewiesen. Als Grundlage für den Generierungsprozeß wird jedoch die Gesamtzahl der Arbeitnehmer verwandt, da die Angaben zu den Arbeitnehmern mit Löhnen und Gehältern in den einzelnen Größenklassen aus Geheimhaltungsgründen sehr lückenhaft sind. Der hierbei gemachte Fehler bei der Berechnung der durchschnittlichen Lohn- bzw. Gehaltssumme ist gering. Für das "Verarbeitende Gewerbe" werden insgesamt 8.209.833 Arbeitnehmer ausgewiesen, von denen 8.162.445 Lohn und Gehalt beziehen (Abweichung ca. 0,58%). Zudem wird die veröffentlichte durchschnittliche Lohn-/Gehaltssumme pro Arbeitnehmer auf Basis der Gesamtzahl der Arbeitnehmer und nicht auf der Zahl der Arbeitnehmer mit Löhnen und Gehältern berechnet (vgl. ebenda, S. 287).

Um einen repräsentativen Datensatz erzeugen zu können, wird nicht nur auf die nach Wirtschaftszweig und Betriebsgrößenklasse differenzierte Tabellierung der Arbeitnehmer zurückgegriffen, sondern auch auf die in gleicher Differenzierung vorliegende Zahl der Unternehmen. Auch bei der Zahl der Unternehmen wird nach solchen unterschieden, die Arbeitnehmer mit Löhnen und Gehältern beschäftigen oder nicht. Insbesondere in den unteren Größenklassen (bis 4 Arbeitnehmer) weicht die Zahl der Unternehmen insgesamt von der Zahl der Unternehmen mit Löhnen und Gehältern erheblich ab. In kleineren Unternehmen ist die Wahrscheinlichkeit groß, daß sie nur aus tätigen Inhabern und unbezahlt mithelfenden Familienangehörigen bestehen und daher keine Arbeitnehmer mit Löhnen und Gehältern beschäftigt sind. Für den Generierungsprozeß wird die Zahl der Unternehmen mit Löhnen und Gehältern verwendet. Die fehlenden

Werte<sup>123</sup> werden so geschätzt, daß die ausgewiesene Gesamtsumme mit der Summe der Einzelwerte pro Größenklasse übereinstimmt. Ein Verzicht auf eine dieser Informationsquellen würde entweder nur zu einem betriebs- oder beschäftigtenproportionalen Datensatz führen.

Die Zahl der Arbeitnehmer und Unternehmen wird auf Unternehmensebene differenziert für 11 Betriebsgrößenklassen und für die 196 Gruppen (Wirtschaftszweige) ausgewiesen.<sup>124</sup>

### **Arbeitnehmer (weiblich)**

Um die von Frick berechneten Segregationsindizes rekonstruieren zu können, muß die Zahl der weiblichen Arbeitnehmer pro Unternehmen generiert werden. Bezieht man diese Variable in die Bestimmung der betrieblichen Schwerbehindertenquote mit ein, dann kann diese Größe in der Regressionsschätzung kontrolliert werden. Es ist zu erwarten, daß mit zunehmendem Frauenanteil die Schwerbehindertenquote geringer ist, da Frauen bei den Schwerbehinderten stark unterrepräsentiert sind (vgl. Frick 1992, S. 176, Frick/Sadowski 1988, S. 25).<sup>125</sup>

Die Zahl der weiblichen Arbeitnehmer wird nicht mehr differenziert nach Betriebsgrößenklasse und Wirtschaftszweig, sondern nur für die einzelnen Wirtschaftszweige ausgewiesen. Eine Verknüpfung zu den anderen Aggregatinformationen erfolgt daher nur über die Variable "Wirtschaftszweig".

### **Facharbeiter oder Gesellen, Ausländer, Auszubildende**

Die Variable "Facharbeiter oder Gesellen" dient als Operationalisierung der Größe "Humankapital" (vgl. Frick 1992, S. 219). Als Facharbeiter oder Gesellen gelten in der Arbeitsstättenzählung die Arbeitnehmer, die "der Arbeiterrentenversicherung unterliegen oder auf Grund besonderer Vorschriften von dieser befreit sind und wegen ihrer Fachkenntnisse und Fähigkeiten mit Arbeiten beschäf-

<sup>123</sup> Aus Geheimhaltungsgründen werden die Zahl der Unternehmen mit Löhnen und Gehältern dann nicht veröffentlicht, wenn aufgrund der Veröffentlichung einzelne Unternehmen identifizierbar wären, d.h. wenn die Zahl der Unternehmen ohne Löhne und Gehälter in der jeweiligen Größenklasse sehr niedrig ist. Geheimzuhalten (in bezug auf Löhne und Gehälter) sind alle Angaben, denen eine Fallzahl von weniger als drei zugrunde lag, sowie Angaben, bei denen der Gesamtwert durch den Anteil eines einzelnen Auskunftspflichtigen dominiert wurde (vgl. Statistisches Bundesamt 1989, Fachserie 2, Heft 1, S. 23).

<sup>124</sup> Die Betriebsgrößenklassen sind wie folgt gegliedert: 1, 2-4, 5-9, 10-19, 20-49, 50-99, 100-199, 200-499, 500-999, 1.000-4.999, 5.000 u. m. Beschäftigte.

<sup>125</sup> Ein hoher Frauenanteil könnte jedoch auch für ein größeres Maß an Organisationskapital stehen, wenn die Nichtbeschäftigung von Frauen als Diskriminierung angesehen wird, d.h. der Einfluß des Frauenanteils auf die Schwerbehindertenquote könnte nicht eindeutig sein (nicht signifikanter Regressionskoeffizient).

tigt werden, die als besonders schwierig oder verantwortungsvoll gelten" (Statistisches Bundesamt 1989, Fachserie 2, Heft 1, S. 8). Diese Operationalisierung von "Humankapital" geht über die von Frick gewählte hinaus, da Qualifikation (allgemeine u. betriebspezifische) nicht notwendigerweise an eine abgeschlossene Berufsausbildung gebunden ist.

Die Variable "Ausländer" wird benötigt, um die Segregationsindizes berechnen zu können. Als Ausländer gelten "alle Personen ohne deutsche Staatsangehörigkeit einschließlich der Staatenlosen und der Personen mit "ungeklärter" Staatsangehörigkeit, die sich im Bundesgebiet aufhalten. Die in der Bundesrepublik anwesenden nicht deutschen Flüchtlinge, die den Status eines heimatlosen Ausländers oder ausländischen Flüchtlings besitzen, zählen zu den Ausländern. Personen, die sowohl die deutsche als auch eine fremde Staatsangehörigkeit haben, gelten als deutsche Staatsangehörige" (ebenda, S. 8).

Einen weiteren Hinweis über die Humankapitalausstattung der Belegschaft eines Unternehmens gibt die Variable "Auszubildende". Je höher der Anteil an Auszubildenden, desto mehr qualifizierte Arbeitskräfte stehen dem Unternehmen potentiell zukünftig zur Verfügung. Die Variable kann zudem als Indiz für Organisationskapital angesehen werden, da in der Beschäftigung von Auszubildenden ein Bestandteil der langfristigen und rücksichtsvollen Nutzung von Humankapital gesehen werden kann. Die Zahl der Auszubildenden pro Unternehmen wird zudem benötigt, um den Schwerbehindertenanteil richtig berechnen zu können: Die in der amtlichen Statistik ausgewiesenen Schwerbehindertenquoten beziehen sich auf Beschäftigte ohne Auszubildende, d.h. bei der Generierung der beschäftigten Schwerbehinderten muß die Zahl der Auszubildenden von der Gesamtzahl der Beschäftigten pro Unternehmen abgezogen werden.

Die Verteilungen der Variablen "Facharbeiter oder Gesellen, Ausländer und Auszubildende" werden in der Arbeitsstättenzählung nicht als gemeinsame Verteilung nach Wirtschaftszweig und Betriebsgrößenklasse ausgewiesen, sondern nur auf Ebene der Wirtschaftszweige. Es erfolgt außerdem keine Veröffentlichung der Anteilswerte dieser Variablen an der Gesamtzahl der Beschäftigten auf Unternehmensebene, so daß die der Simulation zugrundeliegenden Aggregatinformationen erst berechnet werden müssen. Es wird angenommen, daß die Anteilswerte der drei Variablen auf Unternehmensebene gleich den Anteilswerten auf Arbeitsstättenebene sind. Für die Zahl der weiblichen Beschäftigten lassen sich die Anteilswerte auf Arbeitsstätten- und Unternehmensebene berechnen. Ein Vergleich der Werte zeigt nur geringe Unterschiede zwischen den Anteilswerten der beiden Darstellungseinheiten (Arbeitsstätten bzw. Unternehmen). Tabelle 16 zeigt in den Spalten 2 und 3 die entsprechenden Werte auf der Ebene der Unterabteilungen für das Verarbeitende Gewerbe. Da der Unterschied sehr gering ausfällt, kann auch für die Variablen "Facharbeiter oder Gesellen, Ausländer und

Auszubildende" davon ausgegangen werden, daß die Unterschiede sehr gering sind.<sup>126</sup>

Tabelle 16: Vergleich des Anteils der weiblichen Beschäftigten auf Unternehmens- und Arbeitsstättenebene

Branche	Anteil Frauen	
	Betrieb	Unternehmen
Chem. Industrie	26,0	26,6
Herstellung v. Kunststoff- und Gummiwaren	29,2	29,7
Gew. u. Verarb. v. Steinen u. Erden, Feinkeramik, Glas	20,3	21,2
Metallerz. u. -bearbeitung	13,8	14,1
Stahl-, Maschinen- u. Fahrzeugbau	15,5	15,7
Elektrotechnik, Feinmechanik	35,7	35,7
Holz-, Papier- u. Druckgewerbe	26,4	27,1
Leder-, Textil- u. Bekleidungsind.	63,5	63,5
Ernährungsgew., Tabakverarbeitung	42,0	45,1

Quelle: eigene Berechnungen

### Schwerbehinderte

Die Zahl der schwerbehinderten Beschäftigten pro Unternehmen wird einerseits zur Berechnung der Segregationsindizes benötigt und andererseits stellt sie die abhängige Variable im unten dargestellten Regressionsansatz dar. Die Schwerbehindertenquoten werden auf Ebene des Zweistellers der Bundesanstalt für Arbeit durch Auswertung der Arbeitgeberanzeigen gemäß § 13 Abs. 2 SchwbG berechnet und veröffentlicht. Es erfolgt außerdem eine Veröffentlichung der Schwerbehindertenquoten für einzelne Betriebsgrößenklassen. Die Gliederung der Größenklassen weicht jedoch von der in der Arbeitsstättenzählung gewählten Gliederung ab, so daß im Simulationsprozeß der anderen Gliederung Rechnung getragen werden muß. Die von der Bundesanstalt gewählte Unterteilung des Wirtschaftszweiges stimmt ebenfalls nicht mit der in der Arbeitsstättenzählung verwendeten Gliederung der Unterabteilungen (Zweisteller) überein.

Frick (1992, S.257) zeigt die durchschnittlichen Schwerbehindertenquoten für die von der Bundesanstalt für Arbeit gewählte Gliederung als Durchschnittswert für die Jahre 1982-1986. Diese Werte können mit Hilfe der veröffentlichten Statistiken aus der Arbeitsstättenzählung auf die dort gewählte Gliederung umgerechnet werden, so daß für die neun Unterabteilungen eine Schwerbehindertenquote zur

<sup>126</sup> Die Anteilswerte werden in der Simulation nur als Mittelwerte benutzt, um die die tatsächlichen Werte in den einzelnen Unternehmen streuen.

Verfügung steht. Die Generierung der beschäftigten Schwerbehinderten erfolgt auf Basis dieser eindimensionalen Verteilungen (Branchen und Betriebsgrößenklassen).

### **Lohn**<sup>127</sup>

Um den von Frick (1992, S. 152) postulierten Zusammenhang zwischen Schwerbehindertenquote und durchschnittlichen Pro-Kopf-Einkommen im Unternehmen testen zu können, wird die in der Arbeitsstättenzählung erhobene Größe "Löhne und Gehälter" herangezogen. Erfasst sind alle Bruttolöhne und -gehälter der Arbeitnehmer inklusive aller tariflichen oder frei vereinbarter Zulagen (z.B. Akkord-, Nachtarbeits- und Schmutzzulagen, Mietzuschüsse, Kinderzulagen, Fahrkartenzuschüsse, Essensgeld), Naturalvergütungen, Vergütungen für ausgefallene Arbeitszeit (insbesondere Urlaubsgeld), Zuschüsse des Arbeitgebers zum Krankengeld, Urlaubsbeihilfen, Entschädigungen für nicht gewährten Urlaub, Gratifikationen, Gewinnbeteiligungen, Erfolgsprämien, Erfindergeld sowie vermögenswirksame Leistungen.

Zu den Löhnen und Gehältern gehören dagegen nicht Pflichtbeiträge des Arbeitgebers zur Sozialversicherung, Zahlungen aufgrund des Kindergeldgesetzes, allgemeine soziale Aufwendungen (z.B. für Kantinen, Werkskindergärten, Erholungsheime usw.), Spesenersatz, Bergmannsprämien für Bergleute unter Tage, Ruhegehälter und Betriebspensionen, Auslösungen im Baugewerbe, Umzugskostenvergütungen sowie Vorruhestandsbezüge.

Die angegebenen Werte für Lohn-/Gehaltssumme bzw. Durchschnittslohn pro Arbeitnehmer beziehen sich auf das Jahr 1986. Die ausgewiesenen Durchschnittswerte wurden auf die Gesamtzahl der zum Zeitpunkt der Erhebung beschäftigten Arbeitnehmer bezogen. Es ist daher möglich, daß der Durchschnittslohn zu niedrig ausgewiesen wird, wenn Unternehmen in 1986 keine Arbeitnehmer mit Löhnen und Gehältern beschäftigten. Gleichzeitig ist eine Überschätzung des durchschnittlichen Lohns/Gehalts möglich, wenn zum Zeitpunkt der Erhebung weniger Arbeitnehmer beschäftigt waren als 1986. Görzig (1991, S. 66, vgl. unten) weist für das Jahr 1987 in Betrieben mit 20 und mehr Beschäftigten eine geringfügig höhere Beschäftigtenzahl als 1986 aus (1986: 6.996.900; 1987: 7.005.100). Es kann daher von einer geringfügigen Unterschätzung der Durchschnittslohns/-gehalts ausgegangen werden. Eine Analyse der Veränderung in den einzelnen Branchen zeigt, daß die Veränderungsraten der Beschäftigtenzahl sehr gering ausfallen, so daß die Unterschätzung vernachlässigt werden kann.

Die hier verwendete Größe "durchschnittliches Pro-Kopf-Einkommen" geht über die von Frick in seine Analysen einbezogene Größe hinaus. Frick (1992, S. 218) verwendet das um die tariflichen Erhöhungen fortgeschriebene durchschnittliche

---

<sup>127</sup> Vgl. Statistisches Bundesamt, Fachserie 2, Heft 10, S. 7f.

Pro-Kopf-Einkommen der sozialversicherungspflichtig Beschäftigten. "Die Verwendung des Pro-Kopf-Einkommens aus sozialversicherungspflichtiger Beschäftigung stellt aus verschiedenen Gründen eine eher konservative, d.h. die Unterschiede tendenziell unterschätzende Approximation an die tatsächlichen Einkommensdifferenzen dar. Da in Wirtschaftszweigen mit einem niedrigen Durchschnittseinkommen überproportional viele Arbeitnehmer beschäftigt sind, deren Einkommen unterhalb der Pflichtgrenze liegt und deren Verdienst überhaupt nicht in die Berechnung des durchschnittlichen Pro-Kopf-Einkommens eingeht, sind die Durchschnittseinkommen in diesen Wirtschaftszweigen tendenziell zu hoch ausgewiesen. Umgekehrt gibt es in Branchen mit einem hohen Durchschnittseinkommen einen überdurchschnittlichen Anteil an Arbeitnehmern, deren Einkommen oberhalb der Beitragsbemessungsgrenze liegt und die deshalb nur mit einem Einkommen, das exakt der Beitragsbemessungsgrenze entspricht, in die Berechnung des Mittelwertes eingehen. Damit werden die Durchschnittseinkommen in diesen Branchen tendenziell zu niedrig ausgewiesen" (Frick 1992, S. 220).

Neben der Über- bzw. Unterschätzung führt die Verwendung des Pro-Kopf-Einkommens aus sozialversicherungspflichtiger Beschäftigung zu einer Reduktion der Varianz dieser Größe, da das Einkommensspektrum im niedrigen (unterhalb der Pflichtgrenze) und im höheren (oberhalb der Beitragsbemessungsgrenze) Bereich "abgeschnitten" wird.

Der in der Arbeitsstättenzählung erhobene Wert beinhaltet darüber hinaus eine Vielzahl von Komponenten, die über das sozialversicherungspflichtige Arbeitsentgelt hinausgehen.

### **Bruttoanlagevermögen und Bruttowertschöpfung**

In einem vom DIW durchgeführten Forschungsprojekt wurden Zeitreihen von 1977 bis 1988 (mit vorläufigen Schätzungen auch für 1989) für die Beschäftigten, die Wertschöpfung, die Brutto-Anlageinvestitionen und das Brutto-Anlagevermögen ermittelt.<sup>128</sup> Die Werte geben die jeweiligen Jahresanfangsbestände wieder. Erfasst sind alle Betriebe des Verarbeitenden Gewerbes mit 20 und mehr Beschäftigten. Die in dieser Studie verwendete Branchengliederung (Systematik der Wirtschaftszweige, Fassung für das Produzierende Gewerbe) weicht von der in der Arbeitsstättenzählung verwendeten Gliederung der Unterabteilungen und/oder Gruppen ab, so daß eine Anpassung an die Gliederung in der Arbeitsstättenzählung vorgenommen werden muß. Die vom DIW ausgewiesenen Branchen werden mittels der in der Arbeitsstättenzählung ausgewiesenen Untergliederung der Gruppen in die entsprechenden Unterabteilungen umgruppiert. Zur Beurteilung, inwieweit die beiden verwendeten Zahlenwerke ver-

---

<sup>128</sup> Für eine detaillierte Darstellung der Vorgehensweise vgl. Görzig 1991.

gleichbar sind, bzw. die vorgenommene Umgliederung nicht zu erheblichen Verzerrungen führt, kann die Zahl der Beschäftigten in Betrieben mit 20 und mehr Beschäftigten in beiden Zahlenwerken nach der Umgliederung miteinander verglichen werden. Tabelle 17 und Abbildung 8 zeigen die Zahl der Beschäftigten pro Unterabteilung des Verarbeitenden Gewerbes aus der Arbeitsstättenzählung einerseits und der umgegliederten DIW-Statistik andererseits. Auch wenn beiden Erhebungen unterschiedliche Analyseebenen (Unternehmen in der Arbeitsstättenzählung, Betriebe in der DIW-Statistik) zugrundeliegen, zeigen sich nur geringfügige Unterschiede in der Beschäftigtenzahl pro Unterabteilung.<sup>129</sup>

Tabelle 17: Vergleich der Zahl der Beschäftigten aus der Arbeitsstättenzählung und der DIW-Statistik auf der Ebene der Unterabteilungen für das Verarbeitende Gewerbe 1987

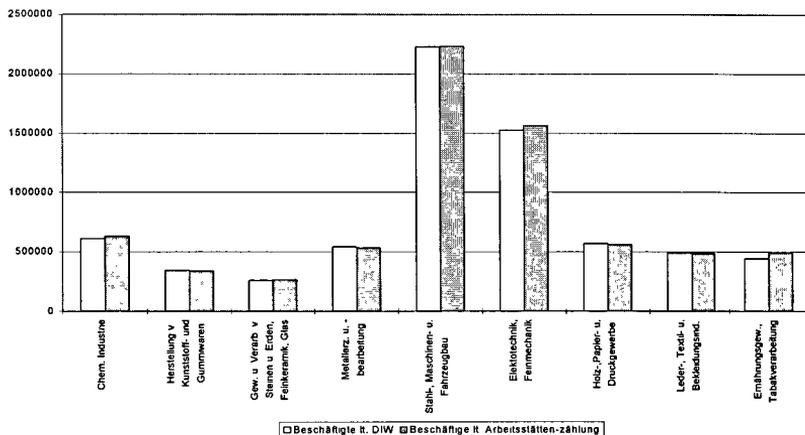
Branche	Beschäftigte lt. DIW	Beschäftigte lt. Arbeitsstättenzählung	Prozentuale Abweichung AZ=100%
Chemische Industrie	609.700	629.022	3,07
Herstellung von Kunststoff- und Gummiwaren	341.600	335.387	-1,85
Gewinnung und Verarbeitung von Steinen und Erden, Feinkeramik, Glas	258.900	263.033	1,57
Metallerzeugung und -bearbeitung	542.400	531.759	-2,00
Stahl-, Maschinen- und Fahrzeugbau	2.229.100	2.231.374	0,10
Elektrotechnik, Feinmechanik	1.524.800	1.562.649	2,42
Holz-, Papier- und Druckgewerbe	567.500	559.759	-1,38
Leder-, Textil- und Bekleidungsindustrie	488.600	484.722	-0,80
Ernährungsgewerbe, Tabakverarbeitung	442.900	495.223	10,57
Summe	7.005.500	7.092.928	1,23

Quelle: Statistisches Bundesamt 1991, Fachserie 2, Heft 10; Görzig 1991

Die ausgewählten Variablen lassen zwar keine vollständige Rekonstruktion der von Frick ermittelten empirischen Resultate zu, aber durch die Bereitstellung der Variablen "Bruttoanlagevermögen" und "Bruttowertschöpfung" können Aussagen über den Zusammenhang von Organisationskapital und Arbeitsproduktivität getroffen werden. Durch die Berücksichtigung des Bruttoanlagevermögens kann sogar der Einfluß einer besseren Kapitalausstattung kontrolliert werden.

<sup>129</sup> Die Beschäftigtenzahl ist die einzige Variable, die zur Überprüfung des Umgliederungsfehlers herangezogen werden kann, da nur diese Variable in beiden Datensätzen erfaßt ist.

Abbildung 8: Vergleich der Zahl der Beschäftigten aus der Arbeitsstättenzählung und der DIW-Statistik auf der Ebene der Unterabteilungen für das Verarbeitende Gewerbe 1987



Quelle: eigene Erstellung

## 4.2 Implementation des Simulated Annealing zur Generierung einer synthetischen Unternehmensstichprobe

### 4.2.1 Vorbemerkungen

Die von Frick (1992) ermittelten empirischen Resultate basieren auf Branchen- und Individualdaten aller Wirtschaftszweige. In dieser Arbeit erfolgt jedoch eine Beschränkung auf das "Verarbeitende Gewerbe"<sup>130</sup>. Dies hat im wesentlichen pragmatische Gründe:<sup>131</sup>

<sup>130</sup> Möglicherweise kann der Vorwurf erhoben werden, daß hier erneut ein bereits recht gut dokumentierter Bereich der empirischen Forschung zum Gegenstand gewählt wird. Die Auswahl erscheint jedoch gerechtfertigt, da damit ein Vergleich von Ergebnissen auf Basis "echter" Daten mit Ergebnissen auf Basis "synthetischer" Daten möglich wird. Eignen sich synthetische Unternehmensdaten zur empirischen Forschung, dann wäre in der Ausweitung des sehr umfangreichen Erhebungsprogrammes des Verarbeitenden Gewerbes auf andere Branchen ein großer Fortschritt zu sehen, da diese Daten, auch wenn sie aggregiert veröffentlicht werden, als Grundlage zur Disaggregation herangezogen werden könnten.

<sup>131</sup> Das Verarbeitende Gewerbe ist der am besten dokumentierte Wirtschaftszweig in der

- Die Aggregatdaten für Bruttowertschöpfung und Bruttoanlagevermögen differenziert nach Betriebsgrößenklassen sind in dieser Differenzierung nur für das Verarbeitende Gewerbe verfügbar.
- Das Verarbeitende Gewerbe stellt mit 8.209.833 Beschäftigten 1987 immer noch die Branche mit den meisten Beschäftigten dar, d.h. Aussagen über die betriebliche Nachfrage nach Arbeitskräften in dieser Branche sind von erheblichem Interesse.

Die zur Erzeugung der synthetischen Stichprobe verwendeten Daten sind in den Tabellen 19 - 30 enthalten. Ausgehend von den in Abschnitt III.3.3.1 formulierten "Allgemeinen und problemspezifischen Spezifikationen" werden im folgenden zunächst die gewählten Spezifikationen bei der Implementation des Simulated Annealing vorgestellt. Dabei wird den oben diskutierten theoretischen Ergebnissen sowie den Simulationsergebnissen der Experimente Rechnung getragen.

Die synthetische Stichprobe wird in zwei Schritten erzeugt: Als Ausgangspunkt für die Generierung weiterer Variablen werden zunächst die Variablen "Branche" und "Beschäftigte" erzeugt. Ein weiterer Simulationslauf erzeugt dann die restlichen Variablen in Abhängigkeit von Branchen- und/oder Größenklassenzugehörigkeit. In jedem Iterationsschritt wird ein Fall und pro Fall eine Variable ausgewählt, deren Ausprägung verändert wird.

Die Programmierung des Algorithmus erfolgte in Turbo Pascal 7.0. Alle durchgeführten Auswertungen wurden mit SPSS PC +, Version 4.01, und MS-EXCEL, Version 5.0, durchgeführt. Soweit zur unmittelbaren Nachvollziehbarkeit der Ergebnisse der Simulation und der sich anschließenden Auswertungen Tabellen und Grafiken notwendig sind, werden sie im Text wiedergegeben.

#### **4.2.2 Bestimmung der a priori-Verteilung und des Generierungsmechanismus**

Zunächst werden die Variablen "Branche" und "Beschäftigte pro Unternehmen" generiert (vgl. unten). Die Zahl der Beschäftigten pro Unternehmen dient als Ausgangspunkt für die Erzeugung einiger weiterer Variablen. Die Größen "weibliche Beschäftigte, Facharbeiter, Auszubildende, Ausländer und Schwerbehinderte" können nur aus einem Anteil der Beschäftigten bestehen. Daher bildet die Zahl der Beschäftigten die Obergrenze für die anderen beschäftigten-zahlabhängigen Variablen.

---

amtlichen Statistik. Es kann daher an diesem Beispiel exemplarisch gezeigt werden, welche zusätzlichen Analysen bei einer Disaggregation entsprechend weitreichender, aggregierter Daten mittels Simulated Annealing ermöglicht würden.

## Branche

Die Zahl der Unternehmen in den einzelnen Branchen kann der Amtlichen Statistik entnommen werden. Entsprechend dieser Verteilung wird die Variable "Branche" den Unternehmen der Stichprobe zugewiesen. Dazu werden die Anteilswerte der einzelnen Branchen eingelesen und die Ausprägung wird einer diesem Anteilswert entsprechenden Zahl von Unternehmen in der Stichprobe zugewiesen. Man erhält nach dieser Zuweisung eine Stichprobe von Unternehmen, die exakt der Branchenverteilung in der Grundgesamtheit entspricht.

## Beschäftigte

Aus der amtlichen Statistik ist zudem die Verteilung der Unternehmen nach Betriebsgrößenklassen in der jeweiligen Branche bekannt, so daß diese "empirische Verteilung" zur Erzeugung der Variablen "Beschäftigte" herangezogen werden kann. Die Verteilung der Beschäftigten pro Unternehmen ist extrem linksschief, d.h. es existiert eine große Zahl von Unternehmen mit wenigen Beschäftigten, so daß die Verwendung einer Gleichverteilung zu extrem niedrigen Anpassungsgeschwindigkeiten führen würde. Die Wahl einer derartigen Zuweisung ist jedoch nur mit den theoretischen Konvergenzeigenschaften vereinbar, weil in dem von Gyárfás/Quinke (1991) vorgestellten Beweis der Konvergenz des Simulated Annealing auf die Symmetrieannahme verzichtet werden kann (vgl. Abschnitt III.3.3.2). Die hier gewählte Nachbarschaftsstruktur ist nicht symmetrisch, d.h.  $g_{xy} \neq g_{yx}$ . Die Wahrscheinlichkeit, einem Unternehmen mit zwei Beschäftigten 1.000 Beschäftigte zuzuweisen, ist wesentlich geringer als der umgekehrte Fall (vgl. Tabelle 18).

Ausgehend von der in Tabelle 18 dargestellten empirischen Verteilungsfunktion wird eine aus einer Gleichverteilung gezogene Zufallszahl mit den den Klassengrenzen der jeweiligen Betriebsgrößenklasse entsprechenden Werten der empirischen Verteilungsfunktion verglichen. Aus der so festgelegten Größenklasse wird dann ein Wert mittels gleichverteilter Zufallszahlen aus diesem Intervall ermittelt und dem zufällig ausgewählten Unternehmen als neue Beschäftigtenzahl zugewiesen<sup>132</sup>.

Diese Verteilung wird als a priori-Verteilung genutzt, mittels der die Variable "Beschäftigte" während des Iterationsprozesses verändert wird.

## Frauen, Facharbeiter, Auszubildende, Ausländer, Schwerbehinderte

Die Anteilswerte dieser Größen sind für die einzelnen Branchen in den Ergebnissen der Arbeitsstättenzählung veröffentlicht.

---

<sup>132</sup> Man beachte, daß auch ein identischer Wert möglich ist  $g_{xx} \neq 0$ .

Ausgehend von der Beschäftigtenzahl pro Unternehmen werden potentielle Nachfolgestände erzeugt, indem pro Unternehmen ein Anteilswert der jeweiligen Variable zugelassen wird, der maximal doppelt so groß sein kann wie der Anteilswert in der Grundgesamtheit, falls der Anteilswert in der Grundgesamtheit kleiner als 0,5 ist. In allen anderen Fällen kann der Anteilswert zwischen Null und eins liegen<sup>133</sup>. Für die Zahl der Schwerbehinderten wird eine Bandbreite vom dreifachen des Anteilswertes zugelassen. Durch diese Vorgehensweise ist sichergestellt, daß für die einzelnen Größen keine Werte vorgeschlagen werden, die größer als die Beschäftigtenzahl sind. Die Schwerbehindertenquoten in den Branchen sind so klein, daß auch nach einer Multiplikation mit drei keine Quote erzeugt wird, die annähernd an eins liegt bzw. größer als eins wird.

Tabelle 18: Relative Häufigkeiten und empirische Verteilungsfunktion der Unternehmen nach Betriebsgrößenklassen und Branchen

	Beschäftigtengrößenklasse										
	1	2-4	5-9	10-19	20-49	50-99	100-199	200-499	500-999	1.000-4.999	5.000 u.m.
Chemische Industrie	1,81	20,77	19,31	17,21	15,20	9,04	6,45	5,46	2,07	2,25	0,43
Verteilungsfunktion	1,81	22,58	41,89	59,10	74,30	83,34	89,79	95,25	97,32	99,57	100,00
Kunststoff- und Gummiwaren	1,21	20,45	22,93	20,75	18,19	7,87	4,50	2,66	0,85	0,52	0,07
Verteilungsfunktion	1,21	21,66	44,59	65,34	83,53	91,40	95,90	98,56	99,41	99,93	100,00
Steine und Erde, Feinkeramik, Glas	1,65	32,25	28,28	17,33	11,74	4,72	1,94	1,39	0,40	0,26	0,04
Verteilungsfunktion	1,65	33,90	62,18	79,51	91,25	95,97	97,91	99,30	99,70	99,96	100,00
Metallerzeugung und -bearbeitung	1,72	39,09	29,46	15,07	8,43	2,77	1,67	1,15	0,32	0,27	0,05
Verteilungsfunktion	1,72	40,81	70,27	85,34	93,77	96,54	98,21	99,36	99,68	99,95	100,00
Stahl-, Maschinen- und Fahrzeugbau	1,15	28,77	29,49	18,71	12,19	4,43	2,49	1,62	0,59	0,46	0,10
Verteilungsfunktion	1,15	29,92	59,41	78,12	90,31	94,74	97,23	98,85	99,44	99,90	100,00
Elektrotechnik, Feinmechanik	1,77	33,85	26,57	16,97	11,27	4,16	2,60	1,75	0,57	0,42	0,07
Verteilungsfunktion	1,77	35,62	62,19	79,16	90,43	94,59	97,19	98,94	99,51	99,93	100,00
Holz-, Papier- und Druckgewerbe	1,53	37,12	30,71	16,30	8,82	2,80	1,44	0,94	0,23	0,11	0,00
Verteilungsfunktion	1,53	38,65	69,36	85,66	94,48	97,28	98,72	99,66	99,89	100,00	100,00
Leder-, Textil- und Bekleidungsindustrie	4,54	44,36	20,13	11,61	9,92	4,58	2,61	1,61	0,44	0,19	0,01
Verteilungsfunktion	4,54	48,90	69,03	80,64	90,56	95,14	97,75	99,36	99,80	99,99	100,00
Ernährungsgewerbe, Tabakverarbeitung	0,55	26,21	40,99	21,04	7,34	1,90	1,08	0,64	0,17	0,08	0,00
Verteilungsfunktion	0,55	26,76	67,75	88,79	96,13	98,03	99,11	99,75	99,92	100,00	100,00

Quelle: eigene Berechnungen aus Angaben des Statistischen Bundesamtes

<sup>133</sup> Diese Vorgehensweise wird an folgendem Beispiel verdeutlicht: Die Zahl der Beschäftigten betrage in dem ausgewählten Unternehmen 1.000. Der Anteilswert der Frauen für die Branche, der das Unternehmen, entstammt betrage 0,4. Dann wird die Zahl der weiblichen Arbeitnehmer bestimmt, indem eine gleichverteilte Zufallszahl aus dem Intervall [0;0,8] gezogen wird. Um die Zahl der weiblichen Beschäftigten zu bestimmen, wird diese Zufallszahl mit der Zahl der Beschäftigten multipliziert und auf einen ganzzahligen Wert gerundet.

## **Durchschnittslohn, Bruttowertschöpfung und Bruttoanlagevermögen pro Arbeitnehmer im Unternehmen**

In den veröffentlichten Ergebnissen der Arbeitsstättenzählung wird der Durchschnittslohn pro Arbeitnehmer nach Branche und Betriebsgrößenklasse ausgewiesen. Zur Konstruktion der Variablen "Durchschnittslohn pro Arbeitnehmer im Unternehmen" wird vom Durchschnittslohn eines Arbeitnehmers in der betreffenden Branche und der betreffenden Betriebsgrößenklasse ausgegangen<sup>134</sup>. Zunächst wird die Branchen- und Größenklassenzugehörigkeit des zufällig im jeweiligen Iterationsschritt ausgewählten Unternehmens bestimmt. Ausgehend von diesen Merkmalen wird dann die durchschnittliche Lohnhöhe generiert, wobei eine Schwankung von 30% um den in der amtlichen Statistik ausgewiesenen Mittelwert für die jeweilige Zelle (Branche und Betriebsgrößenklasse) zugelassen wird.

Bei der Generierung des Bruttoanlagevermögens und der Bruttowertschöpfung pro Arbeitnehmer wird analog verfahren. Der einzige Unterschied besteht darin, daß die Daten des DIW für die Größenklasse von 1.000 und mehr Beschäftigten nicht weiter differenziert sind. Die durchschnittliche Lohnhöhe pro Arbeitnehmer ist noch für Unternehmen mit 5.000 und mehr Beschäftigten in der Arbeitsstättenzählung ausgewiesen.

### **4.2.3 Spezifikation der übrigen Modellparameter**

Wie die Ergebnisse der theoretischen Analyse und die umfangreichen Simulationsexperimente zeigen, nimmt die Wahrscheinlichkeit, einen entropiemaximalen Zustand zu erreichen, zu, wenn mit einer niedrigen Temperaturabsenkungsrate und langen Markovketten gearbeitet wird. Zur Erzeugung der synthetischen Stichprobe wird daher mit folgenden Modellparametern gearbeitet:

Länge der Markovkette:  $5n^{135}$  ( $n=4.300$ )

Temperaturabsenkungsrate: 0,95

Startwert Temperatur: hängt vom anfänglichen Wert der Kostenfunktion ab.

Der Startwert wird so gewählt, daß zu Beginn der Simulation etwa 80% der vorgeschlagenen Verschlechterungen akzeptiert werden. Der Stichprobenumfang wird wegen der Speicherbeschränkungen des verwendeten Personalcomputers auf

---

<sup>134</sup> Die durchschnittliche Lohnhöhe wird in den Veröffentlichungen des Statistischen Bundesamtes als zweidimensionale Verteilung nach Betriebsgrößenklasse (klassifiziert) und Branche ausgewiesen.

<sup>135</sup> Dies bedeutet, daß der Algorithmus für  $5n$  Iterationen bei einer konstanten Temperatur durchgeführt wird. Nach  $5n$  Iterationen wird die Temperatur abgesenkt.

4.300 Unternehmen festgelegt. Die theoretischen Ergebnisse bezüglich der Konvergenz zum Entropiemaximum gelten für große Stichprobenumfänge. In den Simulationsexperimenten wurde gezeigt, daß der hier verwendete Stichprobenumfang von 4.300 als ausreichend angesehen werden kann.

Die Abweichung zwischen den Randverteilungen auf Basis der synthetischen Stichprobe und den vorgegebenen Randverteilungen (mit Index "opt" versehene Größen) wird durch eine additive Kostenfunktion gemessen. Sie hat folgende Form:<sup>136</sup>

$$\begin{aligned}
 C(\text{Tab}, \text{Opt}) = & \sum_{i=1}^9 \sum_{j=5}^{11} \frac{(\text{Zahl d. Betr.}_{ij}^{\text{opt}} - \text{Zahl d. Betr.}_{ij})^2}{\text{Zahl d. Betr.}_{ij}} \\
 & + \sum_{i=1}^9 \sum_{j=5}^{11} \frac{(\text{Beschaeftigte}_{ij}^{\text{opt}} - \text{Beschaeftigte}_{ij})^2}{\text{Beschaeftigte}_{ij}} \\
 & + \sum_{i=1}^9 \sum_{j=5}^{11} \frac{(\text{Lohn}_{ij}^{\text{opt}} - \text{Lohn}_{ij})^2}{\text{Lohn}_{ij}} + \sum_{i=1}^9 \sum_{j=5}^{10} \frac{(\text{Anlage}_{ij}^{\text{opt}} - \text{Anlage}_{ij})^2}{\text{Anlage}_{ij}} \\
 & + \sum_{i=1}^9 \sum_{j=5}^{10} \frac{(\text{Wert}_{ij}^{\text{opt}} - \text{Wert}_{ij})^2}{\text{Wert}_{ij}} + \sum_{i=1}^9 \frac{(\text{Frauen}_i^{\text{opt}} - \text{Frauen}_i)^2}{\text{Frauen}_i} \\
 & + \sum_{i=1}^9 \frac{(\text{Facharbeiter}_i^{\text{opt}} - \text{Facharbeiter}_i)^2}{\text{Facharbeiter}_i} + \sum_{i=1}^9 \frac{(\text{Azubis}_i^{\text{opt}} - \text{Azubis}_i)^2}{\text{Azubis}_i} \\
 & + \sum_{i=1}^9 \frac{(\text{Auslaender}_i^{\text{opt}} - \text{Auslaender}_i)^2}{\text{Auslaender}_i} + \sum_{i=1}^9 \frac{(\text{Behind.}_i^{\text{opt}} - \text{Behind.}_i)^2}{\text{Behinderte}_i} \\
 & + \sum_{l=1}^9 \frac{(\text{Behind. Groessenklasse}_l^{\text{opt}} - \text{Behind. Groessenklasse}_l)^2}{\text{Behinderte Groessenklasse}_l}.
 \end{aligned}$$

Der Index  $i$  steht dabei für die Branche innerhalb des Verarbeitenden Gewerbes, der Index  $j$  für die Betriebsgrößenklasse und der Index  $l$  für die schwerbehindertenspezifische Größenklasse (vgl. Frick 1992, S. 253). Die Werte für das Bruttoanlagevermögen und die Bruttowertschöpfung sind nur für 10 Betriebsgrößenklassen veröffentlicht.

Um eine Abhängigkeit der Ergebnisse vom Startwert des Zufallszahlengenerators auszuschließen, wird die synthetische Stichprobe mit unterschiedlichen Startwer-

<sup>136</sup> Die Variablen entsprechen (auch in der Reihenfolge) den in den Tabellen 19-24 angegebenen Größen. Der letzte Term steht für die größenklassenspezifischen Schwerbehindertenquoten.

ten generiert. Die Analysen zur Anpassung an die Randverteilungen sowie die Hypothesentests erfolgen auf Basis der einzelnen Stichproben. Erweisen sich die Ergebnisse für unterschiedliche Startwerte des Zufallszahlengenerators als konsistent, dann kann der Einfluß des Startwerts auf die Simulationsergebnisse vernachlässigt werden. Die Tabellen 19 bis 24 zeigen die Anpassungen an die Randverteilungen für den Startwert Null des Zufallszahlengenerators. Auch für die anderen Startwerte ergeben sich keine größeren Abweichungen. Auf einen Abdruck wurde daher verzichtet. Die Ergebnisse können daher als unabhängig vom Startwert des Zufallszahlengenerators angesehen werden. Die Qualität der Anpassung an die Randverteilungen unterscheidet sich nur unwesentlich für die einzelnen Startwerte.

#### 4.2.4 Darstellung des Programmablaufs

Die folgenden Ausführungen illustrieren anhand der Darstellung eines Iterationsschritts den Ablauf bei der Generierung der synthetischen Stichprobe.

Zunächst werden die Ausgangsstichprobe (Variable Branche und Zahl der Beschäftigten) und die optimalen Randverteilungen eingelesen, woran sich die Initialisierung der zu generierenden Variablen (Bruttowertschöpfung, Bruttoanlagevermögen, Durchschnittslohn, Facharbeiter, Auszubildende, Frauen, Ausländer und Schwerbehinderte) anschließt. Die Initialisierung erfolgt beispielsweise für die Variable "Bruttowertschöpfung" derart, daß zunächst die Zugehörigkeit des Unternehmens zu einer Branche und Betriebsgrößenklasse bestimmt wird. Gehört das Unternehmen zur Branche "Chemische Industrie" und hat 80 Beschäftigte, dann wird die durchschnittliche Wertschöpfung für dieses Unternehmen durch eine gleichverteilte Zufallszahl aus dem Intervall  $[86.088 - 0,3 \times 86.088; 86.088 + 0,3 \times 86.088]$  bestimmt (vgl. Tabelle 23). Der Wert von 86.088 entspricht der durchschnittlichen Bruttowertschöpfung der Branche "Chemische Industrie" und der Größenklasse 50-99 Beschäftigte. Mit den anderen Variablen wird entsprechend den Ausführungen oben analog verfahren.

An die Initialisierung schließt sich die Bestimmung des Startwerts für den Kontrollparameter an. Der Kontrollparameter wird so bestimmt, daß anfänglich ca. 80% der vorgeschlagenen Verschlechterungen akzeptiert werden. In jedem Iterationsschritt wird durch eine gleichverteilte Zufallszahl aus dem Intervall  $[1; 4.300]$  ein Fall zufällig selektiert. Für diesen Fall wird eine Variable zufällig ausgewählt, deren Wert im aktuellen Iterationsschritt verändert werden soll. Angenommen die zufällige Auswahl führt dazu, daß die Zahl der beschäftigten Facharbeiter verändert werden soll, dann wird zunächst die Branchenzugehörigkeit des selektierten Unternehmens bestimmt. Es sei angenommen, daß das ausgewählte Unternehmen zur "Chemischen Industrie" gehört. Die neue Facharbeiterzahl wird durch Multiplikation der Beschäftigten - und entsprechende Run-

dung auf eine ganze Zahl - mit einer gleichverteilten Zufallszahl aus dem Intervall  $[0; 2 \times 0,2495]$  ermittelt (vgl. Tabelle 24). Nach Zuweisung der neuen Variablenausprägung berechnet das Programm den neuen Wert der Kostenfunktion. Ist der Kostenfunktionswert kleiner als der Wert nach der Initialisierung, wird die Veränderung akzeptiert. Im Falle einer Verschlechterung (der neue Kostenfunktionswert ist größer) wird die Akzeptanzwahrscheinlichkeit anhand der Spezifikation in Gleichung (13) (vgl. Abschnitt III.3.3.1) berechnet. Die so berechnete Akzeptanzwahrscheinlichkeit wird mit einer aus der Menge der auf dem Intervall  $[0; 1]$  gleichverteilten Zufallszahlen gezogenen Zahl verglichen. Ist diese Zahl kleiner oder gleich der Akzeptanzwahrscheinlichkeit, dann wird die Veränderung akzeptiert. Die Wahl des Kontrollparameters stellt dabei sicher, daß die Akzeptanzwahrscheinlichkeit anfänglich im Durchschnitt bei etwa 0,8 liegt, d.h. etwa 80% der vorgeschlagenen Verschlechterungen werden akzeptiert.

Diese Iteration wird für einen konstanten Kontrollparameter 21.500-mal (entspricht dem Fünffachen des Stichprobenumfangs) durchgeführt, bevor der Kontrollparameter durch Multiplikation mit dem Faktor 0,95 abgesenkt wird (äußere Schleife). Im Laufe des Iterationsprozesses erfolgt eine 200-malige Absenkung des Kontrollparameters (entspricht insgesamt 4,3 Mio. Iterationen)<sup>137</sup>.

### 4.3 Ergebnisse der Simulation: Anpassung an die Randverteilungen

Für den Startwert Null des Zufallszahlengenerators werden in den folgenden Tabellen 19 bis 24 die Anpassungen an die vorgegebenen Randverteilungen analysiert. Eine detailliertere Analyse für unterschiedliche Startwerte des Zufallszahlengenerators bestätigt die Ergebnisse für den Startwert Null des Zufallszahlengenerators.

Die Tabelle 19 zeigt die Anpassung der Zahl der Unternehmen nach Branchen und Betriebsgrößenklassen. Die Anpassung gelingt in den einzelnen Zellen sehr gut; es findet sich keine Abweichung, die größer als ein Prozent ist. Größere Abweichungen ergeben sich jedoch in den höheren Betriebsgrößenklassen, da die Zellen der Grundgesamtheit bereits schwach besetzt sind. So sind die Zellen für die Branchen "Holz-, Papier- und Druckgewerbe", "Leder-, Textil- und Bekleidungsindustrie" und "Ernährungsgewinnung, Tabakverarbeitung" in der Größenklasse von 5.000 und mehr Beschäftigten unbesetzt. Da eine betriebs- und beschäftigtenproportionale Stichprobe angestrebt wird, führt ein Auswahlatz von ca. 10% der Unternehmen zu der Unterbesetzung in den erwähnten Fällen. Die Anpassung in den einzelnen Größenklassen über alle Branchen erweist sich wiederum als recht gut. Nur in der Größenklasse von 1.000 bis 4.999 Beschäftigten

---

<sup>137</sup> Bei 4,3 Mio. Iterationen wird ein Unternehmen durchschnittlich in 1.000 Iterationsschritten erfaßt.

wird die Zahl der Unternehmen relativ stark unterschätzt. Insgesamt kann die erzeugte synthetische Stichprobe hinsichtlich der Zahl der Unternehmen nach Branche und Betriebsgrößenklassen als gute Abbildung der Grundgesamtheit betrachtet werden. Die Anpassung der Zahl der Beschäftigten nach Branche und Betriebsgrößenklasse ist Tabelle 20 zu entnehmen. Es zeigt sich, daß die Abweichungen etwas größer als bei der Zahl der Unternehmen ausfallen. Die Anpassung an die vorgegebenen Anteilswerte gelingt jedoch recht gut. Die größten Abweichungen ergeben sich in den oben erwähnten unbesetzten Zellen. Die Anpassung der Zahl der Beschäftigten in den einzelnen Größenklassen über alle Branchen bzw. in den einzelnen Branchen über alle Betriebsgrößenklassen weist nur relativ geringe Differenzen zu den jeweiligen Vorgaben auf. Als ein weiteres Indiz für die Güte der Anpassung kann die Abbildung der Gesamtzahl der Beschäftigten im Verarbeitenden Gewerbe gesehen werden. In der Arbeitsstättenzählung wird für Unternehmen mit 20 und mehr Beschäftigten eine Beschäftigtenzahl von 7.029.928 ausgewiesen; auf Basis der synthetischen Stichprobe ergibt sich nach Hochrechnung mit dem reziproken Wert des Auswahlssatzes eine Beschäftigtenzahl von 7.506.321. Wenn auch größere Abweichungen in einigen Zellen zu beobachten sind, kann die Abbildung bezüglich der Beschäftigtenzahl in der synthetischen Stichprobe trotzdem als mit den Grundgesamtheitswerten verträglich betrachtet werden.

Die Tabellen 21 bis 23 zeigen die Anpassungen für den Lohn, das Bruttoanlagevermögen und die Bruttowertschöpfung pro Beschäftigtem nach Größenklasse und Branche. Die Abweichungen von den vorgegebenen Werten sind gering. Nur bei der Variablen "Durchschnittslohn" ergeben sich aufgrund der unbesetzten Zellen große Abweichungen; die aggregierten Zahlen auf Branchen- bzw. Betriebsgrößenklassenebene weisen jedoch nur kleinere Abweichungen auf. Für die Variable "Bruttoanlagevermögen" und "Bruttowertschöpfung" haben die unbesetzten Zellen eine geringere Bedeutung, da die beiden größten Größenklassen zusammengefaßt sind und die Anpassung in diesen Zellen fast vollständig gelingt. Auch wenn für die Beurteilung der Anpassungsgüte berücksichtigt werden muß, daß die Ursprungsdaten für die Größen "Bruttoanlagevermögen" und "Bruttowertschöpfung" aus einem anderen Datensatz als der Arbeitsstättenzählung stammen und für die Berechnung der entsprechenden Werte auf der Ebene der Branchengliederung in der Arbeitsstättenzählung umfangreiche Aggregationen notwendig waren, kann aufgrund der fast vollständigen Anpassung an die vorgegebenen Werte von einer guten Abbildung der Grundgesamtheit ausgegangen

werden.<sup>138</sup> Für die Variablen "weibliche Beschäftigte, Facharbeiter, Auszubildende, Ausländer und Schwerbehinderte" stehen die Anteilswerte nur auf Branchenebene bzw. für die Schwerbehinderten auch auf Betriebsgrößenklassenebene zur Verfügung. Tabelle 24 zeigt die Anpassungsgüte für diese Variablen. Hier gelingt eine fast vollständige Anpassung an die vorgegebenen Randverteilungen.

Die Analyse der Anpassungsgüte zeigt, daß alle vorgegebenen Randverteilungen durch die auf Basis der synthetischen Stichprobe erzeugten Randverteilungen gut abgebildet werden. Auch unterschiedliche Startwerte des Zufallszahlengenerators haben keinen Einfluß auf die Anpassungsgüte.

Zur Beurteilung der Anpassungsgüte bzw. der Abbildung der Grundgesamtheit kann auch ein Vergleich mit Randinformationen herangezogen werden, die nicht explizit bei der Simulation zugrundegelegt wurden. In den Jahren 1982-1986 beschäftigten durchschnittlich 26,98% der beschäftigungspflichtigen Arbeitgeber im Verarbeitenden Gewerbe keinen Schwerbehinderten. Zur Erzeugung der synthetischen Stichprobe wurde der Durchschnitt der Schwerbehindertenquote dieser Jahre zugrundegelegt. Ein Vergleich der beschäftigungspflichtigen Arbeitgeber gemäß den Angaben der Bundesanstalt für Arbeit mit denen in der synthetischen Stichprobe ermöglicht eine Beurteilung der Abbildung der Grundgesamtheit durch die synthetische Stichprobe, da die Zahl der beschäftigungspflichtigen Arbeitgeber ohne Schwerbehinderte nicht zur Erzeugung der synthetischen Stichprobe herangezogen wurde.

---

<sup>138</sup> Es könnte eingewendet werden, daß die Anpassung an jede einzelne Tabelle nicht bedeutet, daß auch die Variablenbeziehungen entsprechend abgebildet werden. Da aber nach jeder Iteration die Anpassung an alle vorgegebenen Tabellen durch die Kostenfunktion gemessen wird, ist sichergestellt, daß alle in den vorgegebenen Randverteilungen beinhaltenen Strukturinformationen in der synthetischen Stichprobe enthalten sind.

Tabelle 19: Anpassung der Zahl der Unternehmen nach Größenklasse und Branche

Anteilswerte in Prozent der Grundgesamtheit, Unternehmen mit mehr als 20 Beschäftigten

Branche	Größenklasse										Summe Branche
	20 b. 49	50 b. 99	100 b. 199	200 b. 499	500 b. 999	1000 b. 4999	5000 u. m.	Summe Branche			
Chem. Industrie	vorgegeben <i>synth. SP</i>	1.18533 <i>1.23256</i>	0.70528 <i>0.72093</i>	0.50280 <i>0.53488</i>	0.42544 <i>0.47860</i>	0.16153 <i>0.13953</i>	0.17518 <i>0.17628</i>	0.03413 <i>0.02326</i>	3.18969 <i>3.18604</i>		
Herstellung v. Kunststoff- und Gummiwaren	vorgegeben <i>synth. SP</i>	2.91441 <i>2.91023</i>	1.26041 <i>1.30233</i>	0.72121 <i>0.69767</i>	0.42544 <i>0.39535</i>	0.13651 <i>0.13953</i>	0.08418 <i>0.06977</i>	0.01138 <i>0.02326</i>	5.55354 <i>5.55814</i>		
Gew. u. Verarb. v. Steinen u. Erden, Feinkeramik, Glas	vorgegeben <i>synth. SP</i>	2.90531 <i>2.90293</i>	1.16940 <i>1.18605</i>	0.48005 <i>0.46512</i>	0.34354 <i>0.32548</i>	0.09783 <i>0.09302</i>	0.06825 <i>0.04651</i>	0.00910 <i>0.02326</i>	5.07348 <i>5.06977</i>		
Metallverz. u. -bearbeitung	vorgegeben <i>synth. SP</i>	3.96778 <i>3.93569</i>	1.30156 <i>1.22538</i>	0.78491 <i>0.81595</i>	0.54148 <i>0.57163</i>	0.15243 <i>0.13953</i>	0.12741 <i>0.13953</i>	0.02275 <i>0.02326</i>	6.89812 <i>6.80697</i>		
Stahl-, Maschinen- u. Fahrzeugbau	vorgegeben <i>synth. SP</i>	13.49820 <i>13.60465</i>	4.90513 <i>4.90698</i>	2.75743 <i>2.74419</i>	1.79733 <i>1.83721</i>	0.64841 <i>0.67442</i>	0.50507 <i>0.32558</i>	0.10010 <i>0.17628</i>	24.21167 <i>24.20931</i>		
Elektrotechnik, Feinmechanik	vorgegeben <i>synth. SP</i>	9.56454 <i>9.74419</i>	3.53551 <i>3.48837</i>	2.20458 <i>2.30233</i>	1.48564 <i>1.48837</i>	0.48687 <i>0.41886</i>	0.36629 <i>0.18605</i>	0.05688 <i>0.04651</i>	17.70031 <i>17.69768</i>		
Holz-, Papier- u. Druckgewerbe	vorgegeben <i>synth. SP</i>	8.64540 <i>8.72093</i>	2.74833 <i>2.69767</i>	1.41057 <i>1.44786</i>	0.91914 <i>0.93540</i>	0.22296 <i>0.18605</i>	0.11148 <i>0.06977</i>	0.00228 <i>0.00000</i>	14.06016 <i>14.06977</i>		
Leder-, Textil- u. Bekleidungsind.	vorgegeben <i>synth. SP</i>	5.10079 <i>5.13953</i>	2.35701 <i>2.29601</i>	1.34004 <i>1.37209</i>	0.82814 <i>0.79070</i>	0.22751 <i>0.20930</i>	0.09783 <i>0.04651</i>	0.00455 <i>0.00000</i>	9.95587 <i>9.92548</i>		
Ernährungsgew., Tabakverarbeitung	vorgegeben <i>synth. SP</i>	8.75233 <i>8.72093</i>	2.26601 <i>2.25581</i>	1.28543 <i>1.32558</i>	0.76671 <i>0.79070</i>	0.18428 <i>0.18605</i>	0.10010 <i>0.06977</i>	0.00228 <i>0.00000</i>	13.35714 <i>13.34884</i>		
Summe pro Größenklasse	vorgegeben <i>synth. SP</i>	56,53409 <i>56,97674</i>	20,24844 <i>20,27907</i>	11,48702 <i>11,69767</i>	7,53286 <i>7,51163</i>	2,31833 <i>2,20929</i>	1,63579 <i>1,06977</i>	0,24345 <i>0,25583</i>	100 <i>100</i>		

Tabelle 20: Anpassung der Zahl der Beschäftigten nach Größenklasse und Branche

Anteilswerte in Prozent der Grundgesamtheit, Unternehmen mit mehr als 20 Beschäftigten

Branche	Größenklasse										Summe Branche
	20 b. 49	50 b. 99	100 b. 199	200 b. 499	500 b. 999	1000 b. 4999	5000 u. m.				
Chem. Industrie	vorgegeben <i>synth. SP</i>	0,29990 <i>0,28408</i>	0,43738 <i>0,41098</i>	0,79386 <i>0,75333</i>	0,69946 <i>0,66046</i>	2,14959 <i>2,03407</i>	4,25827 <i>6,06286</i>	8,86629 <i>10,42405</i>			
Herstellung v. Kunststoff- und Gummiwaren	vorgegeben <i>synth. SP</i>	0,54716 <i>0,51979</i>	0,62285 <i>0,58856</i>	0,80886 <i>0,76259</i>	0,60419 <i>0,57276</i>	0,95375 <i>0,94452</i>	0,64267 <i>0,68974</i>	4,72846 <i>4,59693</i>			
Gew. u. Verarb. v. Steinen u. Erden, Feinkeramik, Glas	vorgegeben <i>synth. SP</i>	0,53009 <i>0,50222</i>	0,42525 <i>0,40050</i>	0,65001 <i>0,61348</i>	0,41350 <i>0,38265</i>	0,82488 <i>0,77253</i>	0,36607 <i>0,77321</i>	3,70839 <i>3,92312</i>			
Metallerz. u. -bearbeitung	vorgegeben <i>synth. SP</i>	0,56006 <i>0,67925</i>	0,67535 <i>0,63800</i>	1,04188 <i>0,98320</i>	0,63373 <i>0,60394</i>	1,72215 <i>1,61941</i>	2,14446 <i>4,88725</i>	7,49702 <i>9,94049</i>			
Stahl-, Maschinen- u. Fahrzeugbau	vorgegeben <i>synth. SP</i>	2,99892 <i>1,98668</i>	3,41097 <i>2,26789</i>	3,41097 <i>3,24226</i>	2,79419 <i>2,64169</i>	6,07578 <i>5,72732</i>	12,17448 <i>17,78786</i>	31,45313 <i>30,00327</i>			
Elektrotechnik, Feinmechanik	vorgegeben <i>synth. SP</i>	1,76770 <i>1,67034</i>	1,51750 <i>1,43337</i>	1,92607 <i>1,82450</i>	2,10161 <i>1,98709</i>	4,07148 <i>3,84291</i>	7,80487 <i>7,51151</i>	22,03107 <i>20,95807</i>			
Holz-, Papier- u. Druckgewerbe	vorgegeben <i>synth. SP</i>	1,56782 <i>1,48433</i>	1,22411 <i>1,15600</i>	1,68953 <i>1,59776</i>	0,91615 <i>0,86717</i>	1,21665 <i>1,15674</i>	0,12044 <i>0,00000</i>	7,89180 <i>7,35789</i>			
Leder-, Textil- u. Bekleidungsind	vorgegeben <i>synth. SP</i>	0,96833 <i>0,91320</i>	1,00616 <i>0,93337</i>	1,14656 <i>1,08615</i>	0,98894 <i>0,93226</i>	0,98578 <i>0,93429</i>	0,16433 <i>0,00000</i>	6,83386 <i>6,30620</i>			
Ernährungsw., Tabakverarbeitung	vorgegeben <i>synth. SP</i>	1,51925 <i>1,42653</i>	0,94973 <i>0,89563</i>	1,45173 <i>1,02977</i>	0,78722 <i>0,74175</i>	1,10590 <i>1,01670</i>	0,07384 <i>0,00000</i>	6,98193 <i>6,99005</i>			
Summe pro Größenklasse	vorgegeben <i>synth. SP</i>	10,35810 <i>9,79070</i>	8,63785 <i>8,17576</i>	9,94308 <i>9,40315</i>	14,26244 <i>13,47511</i>	9,93899 <i>9,39377</i>	19,10506 <i>18,04709</i>	27,75443 <i>31,77243</i>	100 100		

Tabelle 21: Anpassung für Durchschnittslohn nach Größenklasse und Branche

Unternehmen mit mehr als 20 Beschäftigten

Branche		Größenklasse										Summe Branche
		20 b. 49	50 b. 99	100 b. 199	200 b. 499	500 b. 999	1000 b. 4999	5000 u.m.				
Chem. Industrie	vorgegeben	41.056	45.410	43.468	49.529	51.117	54.160	57.842	53.646			
	<i>synth. SP</i>	47.174	45.353	45.474	49.538	51.121	54.138	57.842	54.840			
Herstellung v. Kunststoff- und Gummiwaren	vorgegeben	32.337	34.018	34.624	37.899	39.430	42.437	44.634	37.561			
	<i>synth. SP</i>	32.127	34.053	34.631	37.968	39.404	41.879	46.033	38.694			
Gew. u. Verarb. v. Steinen u. Erden, Feinkeramik, Glas	vorgegeben	37.265	39.658	40.277	40.791	42.320	43.171	38.943	39.012			
	<i>synth. SP</i>	37.271	39.613	40.278	40.862	42.535	43.794	38.900	40.546			
Metallerz. u. -bearbeitung	vorgegeben	35.889	38.933	40.966	41.939	42.524	46.638	49.095	41.722			
	<i>synth. SP</i>	35.916	38.937	40.954	41.991	42.497	46.620	50.525	46.329			
Stahl-, Maschinen- u. Fahrzeugbau	vorgegeben	35.435	38.892	42.176	43.562	44.818	47.363	51.355	44.566			
	<i>synth. SP</i>	35.440	38.839	42.140	43.508	44.818	47.352	51.315	46.375			
Elektrotechnik, Feinmechanik	vorgegeben	34.173	36.114	37.216	39.399	39.940	41.313	46.905	40.325			
	<i>synth. SP</i>	34.173	36.079	37.158	39.420	39.925	41.176	47.030	41.673			
Holz-Papier- u. Druckgewerbe	vorgegeben	35.672	37.316	38.256	40.017	41.553	43.595	50.000	36.348			
	<i>synth. SP</i>	35.675	37.325	38.288	39.983	41.626	43.351	0	39.175			
Leder-, Textil- u. Bekleidungsind.	vorgegeben	25.020	27.258	29.929	31.437	33.243	33.998	30.000	28.884			
	<i>synth. SP</i>	25.012	27.221	29.936	31.375	33.177	34.358	0	30.287			
Ernährungsgew., Tabakverarbeitung	vorgegeben	24.828	31.460	35.323	38.587	42.160	42.630	49.000	29.316			
	<i>synth. SP</i>	24.821	31.417	35.295	38.574	42.118	42.609	0	35.059			
Summe pro Größenklasse	vorgegeben	32.812	35.985	37.990	40.234	41.989	45.143	50.464	40.402			
	<i>synth. SP</i>	32.816	35.958	37.969	40.230	41.987	45.103	51.015	43.407			

Tabelle 22: Anpassung für Bruttoanlagevermögen pro Beschäftigtem nach Größenklasse und Branche

Unternehmen mit mehr als 20 Beschäftigten

Branche	Größenklasse										Summe Branche	
	20 b. 49	50 b. 99	100 b. 199	200 b. 499	500 b. 999	1000 u. mehr						
Chem. Industrie	vorgegeben <i>synth. SP</i>	248.000 247.997	240.244 240.319	224.100 224.183	241.406 241.444	280.629 280.663	234.895 234.941	234.895 234.941	234.895 234.941	234.895 234.941	234.895 234.941	238.303 238.303
Herstellung v. Kunststoff- und Gummiwaren	vorgegeben <i>synth. SP</i>	84.416 84.473	82.521 82.521	85.267 85.248	91.945 91.960	102.450 102.543	107.094 107.170	107.094 107.170	107.094 107.170	107.094 107.170	107.094 107.170	94.262 95.915
Gew. u. Verarb. v. Steinen u. Erden, Feinkeramik, Glas	vorgegeben <i>synth. SP</i>	253.769 253.727	171.970 171.901	164.723 164.593	148.197 148.246	135.948 135.770	151.958 152.392	151.958 152.392	151.958 152.392	151.958 152.392	151.958 152.392	168.637 166.693
Metallerz u. -bearbeitung	vorgegeben <i>synth. SP</i>	82.799 82.811	91.315 91.247	103.818 103.841	129.117 129.165	174.730 174.770	229.412 229.297	229.412 229.297	229.412 229.297	229.412 229.297	229.412 229.297	172.142 169.667
Stahl-, Maschinen- u. Fahrzeugbau	vorgegeben <i>synth. SP</i>	79.347 79.603	75.618 75.606	77.817 77.871	83.317 83.317	88.200 88.212	131.316 131.348	131.316 131.348	131.316 131.348	131.316 131.348	131.316 131.348	107.972 107.972
Elektrotechnik, Feinmechanik	vorgegeben <i>synth. SP</i>	63.509 63.514	69.138 69.160	72.824 72.820	77.076 77.124	80.329 80.359	95.518 95.597	95.518 95.597	95.518 95.597	95.518 95.597	95.518 95.597	82.896 85.434
Holz-, Papier- u. Druckgewerbe	vorgegeben <i>synth. SP</i>	99.134 99.044	107.358 107.356	99.348 99.348	121.714 121.665	171.259 171.273	159.445 159.193	159.445 159.193	159.445 159.193	159.445 159.193	159.445 159.193	121.269 123.206
Leder-, Textil- u. Bekleidungsind.	vorgegeben <i>synth. SP</i>	74.347 74.428	77.688 77.666	91.875 92.006	99.643 99.679	112.381 112.491	120.655 120.794	120.655 120.794	120.655 120.794	120.655 120.794	120.655 120.794	94.126 96.399
Ernährungsgew., Tabakverarbeitung	vorgegeben <i>synth. SP</i>	144.071 144.171	157.871 157.896	144.207 144.207	155.676 155.686	162.281 162.345	141.440 141.573	141.440 141.573	141.440 141.573	141.440 141.573	141.440 141.573	151.592 150.163
Summe pro Größenklasse	vorgegeben <i>synth. SP</i>	103.077 101.949	100.268 100.778	100.263 100.737	110.168 111.427	125.338 124.404	146.939 153.373	146.939 153.373	146.939 153.373	146.939 153.373	146.939 153.373	123.553 130.714

Tabelle 23: Anpassung für Bruttowertschöpfung pro Beschäftigtem nach Größenklasse und Branche

Unternehmen mit mehr als 20 Beschäftigten

Branche	Größenklasse										Summe Branche
	20 b. 49	50 b. 99	100 b. 199	200 b. 499	500 b. 999	1000 u. mehr					
Chem. Industrie	vorgegeben synth. SP	86.400 86.422	86.179 86.088	79.917 79.870	116.133 116.141	136.289 136.247	91.064 91.096	91.064 91.096	91.064 91.096	91.064 91.096	99.147 95.090
Herstellung v. Kunststoff- und Gummwaren	vorgegeben synth. SP	51.234 51.243	55.402 55.431	57.311 57.257	55.688 55.678	65.880 65.967	59.073 59.272	59.073 59.272	59.073 59.272	59.073 59.272	57.714 57.899
Gew. u. Verarb. v. Steinen u. Erden, Feinkeramik, Glas	vorgegeben synth. SP	59.698 59.772	59.773 59.756	55.044 55.015	58.820 58.823	57.298 57.287	57.598 57.816	57.598 57.816	57.598 57.816	57.598 57.816	58.150 58.115
Metallerz. u. -bearbeitung	vorgegeben synth. SP	43.965 44.030	49.554 49.507	56.855 56.895	57.525 57.531	61.216 61.427	62.696 63.107	62.696 63.107	62.696 63.107	62.696 63.107	58.811 60.028
Stahl-, Maschinen- u. Fahrzeugbau	vorgegeben synth. SP	42.373 42.399	47.126 47.121	46.135 46.135	49.793 49.732	52.074 52.077	68.868 68.923	68.868 68.923	68.868 68.923	68.868 68.923	59.189 60.178
Elektrotechnik, Feinmechanik	vorgegeben synth. SP	43.370 43.422	45.238 45.226	47.782 47.844	55.432 55.452	62.504 62.517	64.053 64.055	64.053 64.055	64.053 64.055	64.053 64.055	57.438 58.461
Holz-, Papier- u. Druckgewerbe	vorgegeben synth. SP	45.435 45.526	49.766 49.786	53.103 53.113	58.561 58.542	68.874 68.806	80.225 81.381	80.225 81.381	80.225 81.381	80.225 81.381	57.434 58.588
Leder-, Textil- u. Bekleidungsind.	vorgegeben synth. SP	36.774 26.851	33.237 33.237	39.750 39.731	43.466 43.356	46.603 46.600	62.698 62.469	62.698 62.469	62.698 62.469	62.698 62.469	40.594 42.164
Ernährungs- u. Tabakverarbeitung	vorgegeben synth. SP	69.876 69.903	79.683 79.643	80.157 80.157	95.883 95.890	105.234 105.237	112.704 112.825	112.704 112.825	112.704 112.825	112.704 112.825	90.366 89.172
Summe pro Größenklasse	vorgegeben synth. SP	46.998 48.082	51.444 51.864	53.487 53.389	60.462 60.950	68.098 67.038	70.862 71.068	70.862 71.068	70.862 71.068	70.862 71.068	62.682 63.864

Tabelle 24: Anpassung für Frauen, Facharbeiter, Azubis, Ausländer und Schwerbehinderte nach Branche

Unternehmen mit mehr als 20 Beschäftigten

Branche		Anteilswerte				
		Frauen	Facharbeiter	Azubis	Ausländer	Schwerbeh.
Chem. Industrie	vorgegeben	26,65	25,12	4,85	7,32	6,07
	<i>synth. SP</i>	26,62	24,95	4,85	7,31	6,17
Herstellung v. Kunststoff- und Gummiwaren	vorgegeben	29,66	22,20	3,87	14,15	5,42
	<i>synth. SP</i>	29,66	22,19	3,86	14,16	5,50
Gew. u. Verarb. v. Steinen u. Erden, Feinkeramik, Glas	vorgegeben	21,23	29,88	4,05	10,26	6,18
	<i>synth. SP</i>	21,23	29,91	4,04	10,27	6,28
Metallerz. u. -bearbeitung	vorgegeben	14,14	31,74	6,87	14,64	6,89
	<i>synth. SP</i>	14,15	31,75	6,86	14,62	7,00
Stahl-, Maschinen- u. Fahrzeugbau	vorgegeben	15,65	38,56	7,52	10,42	5,10
	<i>synth. SP</i>	15,65	38,56	7,51	10,42	5,19
Elektrotechnik, Feinmechanik	vorgegeben	35,70	24,65	5,61	10,51	5,54
	<i>synth. SP</i>	35,69	24,65	5,61	10,51	5,63
Holz-, Papier- u. Druckgewerbe	vorgegeben	27,09	35,61	7,31	7,03	5,39
	<i>synth. SP</i>	27,07	35,62	7,30	7,03	5,48
Leder-, Textil- u. Bekleidungsind.	vorgegeben	63,48	29,84	5,93	11,39	5,74
	<i>synth. SP</i>	63,47	29,86	5,92	11,40	5,72
Ernährungsw., Tabakverarbeitung	vorgegeben	45,13	27,13	13,32	6,36	4,87
	<i>synth. SP</i>	45,13	27,13	13,28	6,35	4,94
Summe	vorgegeben	28,52	30,31	6,75	9,81	5,63
	<i>synth. SP</i>	27,48	30,95	6,73	10,25	5,64

Tabelle 25 zeigt die Zahl der beschäftigungspflichtigen Arbeitgeber ohne Schwerbehinderte. Die Werte in den synthetischen Stichproben liegen zwischen 24,5% und 25,2%, das heißt sie nähern sich dem Wert in der Grundgesamtheit an. Die Anpassung gelingt demnach unabhängig vom Startwert des Zufallszahlengenerators sehr gut.

Tabelle 25: Beschäftigungspflichtige Arbeitgeber ohne Schwerbehinderte

Startwert	Anteil Arbeitgeber ohne Schwerbehinderte
0	25,2
1	25,2
2	24,7
3	24,5
4	25,1
5	24,8

Bei den angegebenen Mittelwerten und Anteilswerten, außer für die Variablen "Branche" und "Zahl der Beschäftigten", handelt es sich um Mittelwerte, die durch Gewichtung mit der Zahl der Beschäftigten berechnet wurden. Die Größen gehen auch so in die Kostenfunktion ein. Berechnet man die Mittelwerte auf Unternehmensebene (ungewichtet), so zeigen sich erhebliche Differenzen zu den auf Branchenebene berechneten Werten. Werden die Branchenaggregate, wie in vielen empirischen Studien auf Basis von Aggregatdaten geschehen, als repräsentative Unternehmen aufgefaßt, dann zeigen die hier ermittelten Werte auf Unternehmensebene, daß die durchschnittlichen Werte von den Aggregatwerten erheblich abweichen.

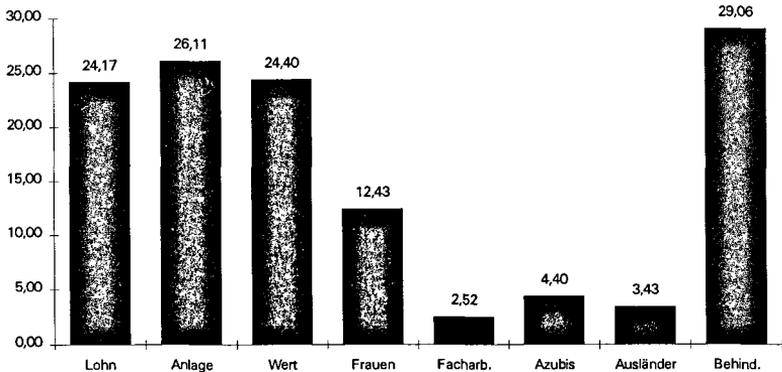
Die Interpretation von Aggregaten als repräsentatives Unternehmen und eine daraus resultierende Verwendung von Aggregatdaten für empirische Studien über Verhaltensweisen von Mikroeinheiten führt zu einer erheblichen Reduktion der Varianz der einzelnen Größen: Da Branchenaggregate zum Beispiel für Frauenanteil und/oder Schwerbehindertenanteil so gut wie nie den Wert Null annehmen, in der Grundgesamtheit sich aber mit hoher Wahrscheinlichkeit Unternehmen mit einer Frauenquote oder einem Schwerbehindertenanteil von Null befinden, wird bei einer Verwendung der Branchenaggregate die Varianz erheblich unterschätzt und der Mittelwert auf Unternehmensebene überschätzt. Frick (1992, S. 254) weist für das Jahr 1987 für das Verarbeitende Gewerbe einen Anteil von 26.2% beschäftigungspflichtiger Arbeitgeber ohne Schwerbehinderte aus, die bei der Arbeit mit den Aggregatdaten ausgeblendet werden. Die Tabelle 26 und die Abbildung 9 zeigen die Abweichungen der auf Basis der Aggregate und der auf Unternehmensebene berechneten Werte für die in der Stichprobe befindlichen Variablen.

Tabelle 26: Mittelwerte auf Unternehmens- und Branchenebene

Variable	Unternehmensebene	Branchenebene	Differenz
			Untern. = 100
Lohn pro Beschäftigtem	34.956,92	43.406,97	24,17
Bruttoanlagevermögen pro Beschäftigtem	103.651,90	130.714,35	26,11
Bruttowertschöpfung pro Beschäftigtem	51.335,44	63.863,62	24,40
Frauenanteil	31,38	27,48	-12,43
Facharbeiteranteil	30,19	30,95	2,52
Anteil Auszubildende	7,04	6,73	-4,40
Anteil Ausländer	9,91	10,25	3,43
Anteil Schwerbehinderte	4,37	5,64	29,06

Infolge dieser bei einigen Variablen erheblichen Abweichungen zwischen den Durchschnittswerten muß bezweifelt werden, ob Aggregatdaten sich für empirische Analysen eignen, d.h. Aggregate als repräsentatives Unternehmen aufzufassen, führt möglicherweise zu falschen Schlußfolgerungen.

Abbildung 9: Absolute Differenz der Mittelwerte auf Unternehmens- und Branchenebene



#### 4.4 Empirische Ergebnisse zu den Bestimmungsgründen der betrieblichen Schwerbehindertenbeschäftigung und deren Einfluß auf die Bruttowertschöpfung auf Basis der synthetischen Stichprobe

##### 4.4.1 Vergleich der Segregationsindizes

Zur Beurteilung, ob Schwerbehinderte hinsichtlich ihrer Beschäftigungschancen im Vergleich zu anderen Beschäftigtengruppen (z.B. Frauen, Ausländer) diskriminiert werden (vgl. Hypothese 1.1), berechnet Frick (1992, S. 187ff) Segregati-

ons- bzw. Dissimilaritätsindizes, die erstmals von Gibbs (1965) zur Untersuchung von beruflicher Segregation verwendet wurden<sup>139</sup>. Trotz umfangreicher Kritik und ebenfalls stark kritisierte Verbesserungsvorschläge kann der Dissimilaritätsindex in seiner ursprünglichen Form nach wie vor als das "Standardmaß" zur zusammenfassenden Kennzeichnung ungleicher Chancenstrukturen zwischen zwei Gruppen bezeichnet werden (vgl. Handl 1984, S. 330, Frick 1992, S. 187).

"Von "Segregation" kann man dann sprechen, wenn schwer- und nichtbehinderte Arbeitnehmer (überwiegend) in unterschiedlichen Berufen und/oder Wirtschaftszweigen beschäftigt sind. "Integration" liegt demgegenüber dann vor, wenn der Schwerbehindertenanteil in jedem Beruf bzw. in jeder Branche (in etwa) dem in der Gesamtwirtschaft entspricht" (Frick 1992, S. 187). Die von Frick auf Branchenebene berechneten Segregationsindizes können mit dem synthetisch erzeugten Datensatz auf Unternehmensebene berechnet werden. Der Index kann interpretiert werden als Anteil der schwer- bzw. nichtbehinderten Arbeitnehmer (auch Frauen und/oder Ausländer), die das Unternehmen wechseln müßten, um eine Gleichverteilung zwischen den Gruppen zu erreichen.

Der Index läßt sich mit folgender Formel berechnen:

$$(36) \quad D = \frac{1}{2} \sum_{i=1}^n \left| \frac{SB_i}{SB} - \frac{NB_i}{NB} \right| * 100$$

wobei     SB:    Zahl der beschäftigten Schwerbehinderten,  
           SB<sub>i</sub>:   Zahl der im i-ten Unternehmen beschäftigten Schwerbehinderten,  
           NB:    Zahl der beschäftigten Nichtbehinderten und  
           NB<sub>i</sub>:   Zahl der im i-ten Unternehmen beschäftigten Nichtbehinderten.

Der Index kann theoretisch Werte zwischen 0 und 100 annehmen. Die Ausprägung 0 bedeutet dabei vollständige Integration und die Ausprägung 100 vollständige Segregation. Ein Vergleich der Segregationsindizes für verschiedene Arbeitnehmergruppen ermöglicht eine Aussage über die relative Segregation der einzelnen Arbeitnehmergruppe im Vergleich zu einer anderen. Frick (1992, S. 190f.) kommt für die Segregation von Behinderten, Frauen und Ausländern zu dem Ergebnis, daß die Segregation Schwerbehinderter bezüglich der sozialrechtlichen Stellung, der beruflichen Stellung und der Branchen im Vergleich zu Frauen und Ausländern wesentlich niedriger ist. Tabelle 27 und Abbildung 10 zeigen die auf Unternehmensebene berechneten Segregationsindizes für unterschiedliche

---

<sup>139</sup> Zur Kritik dieses Index vgl. z. B. Winship 1977, Cortese et. al. 1976. Verbesserungsvorschläge (Taeuber/Taeuber 1976, Falk et. al. 1978, Massey 1978) haben jedoch auch Widerspruch hervorgerufen.

Startwerte des Zufallszahlengenerators sowie den Durchschnittswert über die Stichproben.

Im Unterschied zu den Ergebnissen von Frick<sup>140</sup> weist die Gruppe der Schwerbehinderten annähernd gleich große Indexwerte auf wie die Vergleichsgruppen Frauen und Ausländer. Im Durchschnitt müßten etwa 30% der Schwerbehinderten das Unternehmen wechseln, damit eine Gleichverteilung gegeben wäre. Es kann daher von einer erheblichen Beschäftigungssegregation gesprochen werden. Die doch erhebliche Differenz zu den von Frick auf Basis von Branchendaten (Aggregation von Daten des Sozio-ökonomischen Panels) berechneten Werten kann folgende Ursache haben:

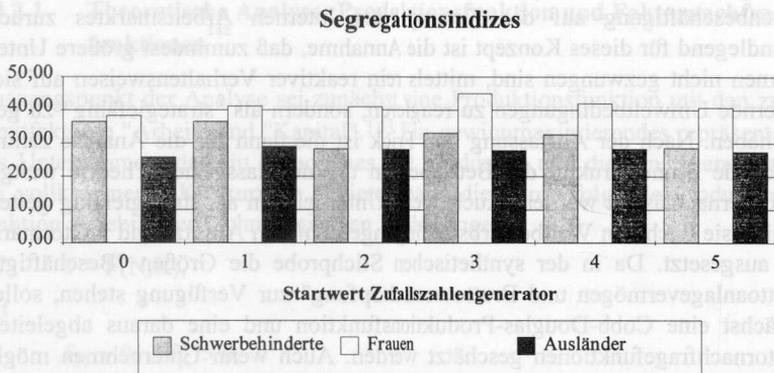
Tabelle 27: Segregationsindizes für Schwerbehinderte, Frauen und Ausländer

Startwert	Schwerbehinderte	Frauen	Ausländer
0	36,49	38,28	25,75
1	36,48	41,18	29,05
2	28,32	41,00	28,14
3	33,08	45,00	27,84
4	32,64	42,02	28,29
5	28,56	40,72	27,45
Durchschnitt	32,60	41,37	27,75
Standardabw.	3,61	2,18	1,12

---

<sup>140</sup> Frick (1992, S.189ff) berechnet die Segregationsindizes für die sozialrechtliche und berufliche Stellung. Eine Berechnung der Indizes für Branchen auf Basis des Zweistellers der Bundesanstalt für Arbeit zeigt, daß die relative Segregation von Schwerbehinderten vergleichsweise gering ausfällt. Nach diesen Berechnungen ist der Anteil der Ausländer und Frauen, der die Branche wechseln müßte, etwa zwei- bzw. viermal so hoch wie bei den Schwerbehinderten.

Abbildung 10: Segregationsindizes für Schwerbehinderte, Frauen und Ausländer



Bei der Verwendung von Branchendaten wird keine Branche mit einem Schwerbehindertenanteil von Null existieren. Die synthetische Stichprobe enthält ca. 25% beschäftigungspflichtiger Betriebe, die keinen einzigen Schwerbehinderten beschäftigen. Aus diesem Grunde erhöht sich der Segregationsindex im Vergleich zu dem von Frick berechneten Wert auf der Ebene der Branchendaten erheblich und gleicht sich den Werten für Frauen und Ausländer an.

Den von Frick berechneten Segregationsindizes liegen die Daten für die Gesamtwirtschaft zugrunde. Die Anteilswerte für Frauen schwanken in der Gesamtwirtschaft in viel stärkerem Maße als im Verarbeitenden Gewerbe. Auf Basis des Zweistellers in der Arbeitsstättenzählung ergeben sich für den Frauenanteil Werte zwischen 0,5% und 95%. Die Anteilswerte im Verarbeitenden Gewerbe weisen dagegen eine sehr viel geringere Bandbreite auf, so daß der auf Branchenebene berechnete Segregationsindex für Frauen in der Gesamtwirtschaft erheblich über dem im Verarbeitende Gewerbe ermittelten Wert liegen muß. Für das Verarbeitende Gewerbe kann daher von einer erheblichen Beschäftigungssegregation von Schwerbehinderten ausgegangen werden, die in ihrer Größenordnung der von Frauen und Ausländern vergleichbar ist.

Ursächlich für den von Frick gefundenen Befund, daß Schwerbehinderte hinsichtlich ihrer Beschäftigungschancen in geringerem Maße diskriminiert werden als Frauen und Ausländer, ist, daß bei der Berechnung der Indizes keine Branchen mit einem Schwerbehindertenanteil von Null berücksichtigt werden. Allein die Tatsache, daß ca. 26% der Unternehmen keinen Schwerbehinderten beschäftigen, zeigt, daß es eine erhebliche Segregation gibt. Auch dieses Ergebnis legt die Schlußfolgerung nahe, daß Aggregatdaten zur Beurteilung von Beschäftigungschancen verschiedener Arbeitnehmergruppen ungeeignet sind.

#### 4.4.2 Bestimmungsgründe der betrieblichen Arbeitsnachfrage: Produktions- und Faktornachfragefunktion

Frick greift in seiner Analyse der Determinanten der betrieblichen Schwerbehindertenbeschäftigung auf das Konzept des Internen Arbeitsmarktes zurück. Grundlegend für dieses Konzept ist die Annahme, daß zumindest größere Unternehmen nicht gezwungen sind, mittels rein reaktiver Verhaltensweisen auf sich ändernde Umweltbedingungen zu reagieren, sondern als "strategiefähig" zu gelten haben. Nach der Auffassung von Frick ist die dann für die Analyse zentral werdende Binnenstruktur der Betriebe von der neoklassischen Theorie weitgehend vernachlässigt worden. Auch wenn Unternehmen als strategiefähig gelten, so sind sie doch den Wettbewerbsbedingungen auf den Absatz- und Faktormärkten ausgesetzt. Da in der synthetischen Stichprobe die Größen "Beschäftigte, Bruttoanlagevermögen und Bruttowertschöpfung" zur Verfügung stehen, sollen zunächst eine Cobb-Douglas-Produktionsfunktion und eine daraus abgeleitete Faktornachfragefunktionen geschätzt werden. Auch wenn Unternehmen möglicherweise strategiefähig sind, ist nicht zu erwarten, daß durch die Beschäftigung bestimmter Arbeitnehmergruppen dauerhaft Wettbewerbsnachteile in Kauf genommen werden. Die Schätzung einer Produktionsfunktion müßte daher einen positiven Zusammenhang zwischen der Beschäftigung Schwerbehinderter (Frauen, Ausländer) und der Bruttowertschöpfung pro Arbeitnehmer liefern, sofern die theoretischen Überlegungen von Frick zutreffen. Dabei muß jedoch der Einfluß anderer Faktoren wie Kapitalintensität, Konzentrationsgrad und Lohnhöhe<sup>141</sup> kontrolliert werden.

Ergeben sich aufgrund der geschätzten Produktionsfunktion signifikante positive Einflüsse der Beschäftigung von Schwerbehinderten (Frauen, Ausländern) auf die Bruttowertschöpfung, kann dies als Indiz für die Gültigkeit der von Frick formulierten theoretischen Überlegungen gewertet werden. Die Analyse der Faktornachfragefunktionen ermöglicht einerseits Aussagen darüber, ob die Bestimmungsgründe der Nachfrage nach diesen Arbeitnehmergruppen von den Bestimmungsgründen der Gesamtnachfrage nach Arbeitnehmern abweichen bzw. ob sich die Bestimmungsgründe für einzelne Arbeitnehmergruppen unterscheiden und andererseits Aussagen darüber, welche Merkmale Unternehmen aufweisen,

---

<sup>141</sup> Der Zusammenhang zwischen Lohnhöhe und Arbeitsproduktivität wird in Effizienzlohnmodellen postuliert (vgl. z.B.: Yellen 1984; Stiglitz 1987; Gerlach/Hübler 1989). Einige Firmen können ihre Gewinne maximieren, wenn sie eine über dem Markträumungslohnsatz liegende Lohnprämie zahlen und die damit verbundenen Produktivitätssteigerungen die Kosten der Lohnprämie kompensieren (vgl. Gerlach/Schmidt 1989, S. 356). In der Literatur sind mehrere Ursachen von Produktivitätssteigerungen modelliert worden: Fluktuationkosten (Salop 1979), Kontroll- oder Shirkingkosten (Shapiro/Stiglitz 1984) sowie eher soziologische Begründungen wie Fairneß, Loyalität und Teamarbeit (Akerlof 1982).

die vermehrt Arbeitnehmer aus den einzelnen Arbeitnehmergruppen beschäftigen.

#### 4.4.2.1 Theoretische Analyse: Produktionsfunktion und Faktornachfragefunktionen<sup>142</sup>

Ausgangspunkt der Analyse sei zunächst eine Produktionsfunktion mit den zwei Inputfaktoren "Arbeit" und "Kapital".<sup>143</sup> Ein gewinnmaximierendes repräsentatives Unternehmen, das ein homogenes Gut produziert und dies an einem Markt bei vollkommener Konkurrenz anbietet, habe die durch folgende Produktionsfunktion abgebildeten technologischen Bedingungen:

$$(37) \quad Y = F(N, K),$$

mit

$$(38) \quad \begin{aligned} f_N &> 0, & f_K &> 0, \\ f_{NN} &< 0, & f_{KK} &< 0, \\ f_{NK} &> 0, \end{aligned}$$

wobei K den Kapitalstock und N den Input homogener Arbeitsleistungen bezeichnet. Dabei bezeichnet  $f$  die Ableitung der Produktionsfunktion hinsichtlich der indizierten Größen.

Gemäß der Annahme einer vollständigen Konkurrenz auf dem Gütermarkt reagiert die Nachfrage nach den Produktionsfaktoren vollkommen preiselastisch, so daß sich folgende Zielfunktion ergibt:

$$(39) \quad F(N, K)p - wN - rK = \max!,$$

mit  $p$  als Outputpreis,  $w$  als Lohn pro Einheit von  $N$  und  $r$  als Kosten für die Nutzung einer Einheit  $K$ . Es ergeben sich demnach folgende Bedingungen erster Ordnung:

<sup>142</sup> Vgl. zu diesem Abschnitt: Hamermesh 1993, S. 22ff, Breitung 1992, S. 143ff, Varian 1981, S. 121ff, Varian 1990, S. 299ff.

<sup>143</sup> Zur Herleitung der Schätzfunktion wird nur der komparativ-statische Fall betrachtet, da es sich bei dem vorliegenden Datensatz um Querschnittsdaten handelt, d.h. es können keine dynamischen Modelle geschätzt werden. Bei Kenntnis entsprechender dynamischer Prozesse können in weitergehenden Simulationen auch Entwicklungsprozesse hinzusimuliert werden, so daß ein disaggregierter Datensatz resultieren würde, der sich für Schätzungen dynamischer Modelle eignet. In der vorliegenden Arbeit wird jedoch zunächst die generelle Eignung synthetischer Daten vor dem Hintergrund einer substanzwissenschaftlichen Fragestellung analysiert, so daß die Beschränkung auf eine komparativ-statische Betrachtungsweise als legitim erscheint.

$$(40) \quad f_N = \frac{w}{p} \text{ und } f_K = \frac{r}{p},$$

d.h. das Grenzprodukt entspricht dem realen Faktorpreis. Annahmegemäß ist die zweite Ableitung der Produktionsfunktion negativ, d.h. es handelt sich um ein Gewinnmaximum. Unterstellt man für die Produktionsfunktion eine Cobb-Douglas-Technologie, dann ergeben sich für die Produktionsfunktion und die Faktornachfragefunktionen folgende Terme:

$$(41) \quad Y = N^\alpha K^\beta,$$

$$(42) \quad N = \alpha \frac{pY}{w}, \text{ bzw. nach Logarithmierung: } \ln N = \text{const} + \ln Y - \ln \frac{w}{p},$$

$$(43) \quad K = \beta \frac{pY}{r}, \text{ bzw. nach Logarithmierung: } \ln K = \text{const} + \ln Y - \ln \frac{r}{p}.$$

Alternativ können die Faktornachfragefunktionen auch aus der Kostenfunktion abgeleitet werden, da unter den oben formulierten Voraussetzungen das Prinzip der Dualität gilt (vgl. Varian 1981, S. 35ff): "Die Kostenfunktion einer Unternehmung beinhaltet alle ökonomisch relevanten Aspekte ihrer Technologie" (Varian 1981, S. 39).

Die Kostenfunktion einer Cobb-Douglas-Technologie hat folgende Gestalt (vgl. Hamermesh 1993, S. 29):

$$(44) \quad C(w, r, Y) = Z w^\alpha r^\beta Y, \text{ mit } Z \text{ als Konstante.}$$

In der logarithmierten Form ergibt sich:

$$(44a) \quad \ln C(w, r, Y) = \ln Z + \alpha \ln w + \beta \ln r + \ln Y.$$

Folglich hängen die Gesamtkosten nur noch vom Outputniveau  $Y$  und den Preisen der Inputfaktoren  $r$  und  $k$  ab. Diese Beziehung eignet sich sehr gut als Schätzfunktion, da die Niveaus der Faktoren Arbeit und Kapital nicht benötigt werden. In der synthetischen Stichprobe sind jedoch sowohl die Zahl der Beschäftigten als auch das Anlagevermögen enthalten, so daß hier die Produktionsfunktion auch direkt geschätzt werden kann. Äquivalent zu einer Schätzung der Produktionsfunktion ist jedoch wegen des Prinzips der "Dualität von Kosten- und Produktionsfunktion" auch die Schätzung der Kostenfunktion, da diese alle ökonomisch relevanten Aspekte der Produktionstechnologie einer Unternehmung beinhaltet.

Mit Shephard's Lemma (vgl. Varian 1981, S. 32; Shephard 1970) angewendet für  $N$  und  $K$  folgt:

$$(45) \quad \frac{N}{K} = \frac{\alpha}{\beta} \frac{r}{w}, \text{ bzw. logarithmiert: } \ln\left(\frac{N}{K}\right) = \alpha' + \ln\left(\frac{r}{w}\right),$$

mit  $\alpha'$  als Konstante. Als Faktornachfragefunktionen erhält man durch Auflösung nach  $\ln K$  bzw.  $\ln N$ :

$$(46) \quad \ln N = \alpha' + \ln r - \ln w + \ln K$$

$$(47) \quad \ln K = \alpha'' - \ln r + \ln w + \ln N.$$

Im hier vorliegenden Fall einer separablen Produktionstechnologie sollten bei der Schätzung der Gleichungen für  $\ln K$  und  $\ln N$  Koeffizienten resultieren, deren Wert nahe bei 1 oder -1 liegt.

Bisher wurde nur der Zweifaktorenfall mit homogener Kapitalausstattung und homogenem Arbeitsinput betrachtet. Der Arbeitsinput kann jedoch mit dem vorliegenden Datensatz hinsichtlich verschiedener Arbeitnehmergruppen (weibliche Arbeitnehmer, Facharbeiter, Auszubildende, Ausländer und Schwerbehinderte) differenziert werden. Es wird unterstellt, daß die Faktornachfrage für eine Arbeitnehmergruppe unabhängig von der Nachfrage nach einer anderen Arbeitnehmergruppe ist. Die Produktionsfunktion aus (41) hat dann folgende Gestalt:

$$(48) \quad Y = N_1^{\alpha_1} N_2^{\alpha_2} K^{\alpha_3} \text{ mit } \sum \alpha_i = 1,$$

bzw. logarithmiert:

$$(49) \quad \ln Y = \alpha_1 \ln N_1 + \alpha_2 \ln N_2 + \alpha_3 \ln K.$$

Hieraus ergeben sich als Bedingungen erster Ordnung für das Gewinnmaximum folgende Beziehungen für  $N_1$  und  $N_2$ :

$$(50) \quad \ln N_1 = \ln \alpha_1 + \ln Y - \ln \frac{w_1}{p}, \quad \ln N_2 = \ln \alpha_2 + \ln Y - \frac{w_2}{p}.$$

$w_1$  und  $w_2$  bezeichnen den Preis für den jeweiligen Arbeitsinput. Die Nachfrage nach einem Produktionsfaktor ist dabei unabhängig von dem Einsatz eines anderen Produktionsfaktors. Sie ist nur vom Output und dem faktorspezifischen Reallohn abhängig. Da bei einer kurzfristigen Betrachtung das Preisniveau als für alle Unternehmen gleich angesehen werden kann, ist die Nachfrage nach Arbeit einer bestimmten Arbeitnehmergruppe nur von deren Nominallohn  $w_i$  abhängig.

#### 4.4.2.2 Schätzung der Produktionsfunktion und Faktornachfragefunktionen<sup>144</sup>

Ausgehend von den Überlegungen in Punkt 4.4.2.1 wird die Schätzgleichung der Produktionsfunktion in einem ersten Schritt wie folgt spezifiziert:

$$(51) \quad \ln Y = \text{const} + \alpha \ln N + \beta \ln K$$

Der Output eines Unternehmens  $Y$  wird durch die Größe Bruttowertschöpfung pro Beschäftigtem gemessen und als abhängige Variable für die Schätzung verwendet. Unter Bruttowertschöpfung versteht man die mit den jeweiligen Verkaufspreisen bewerteten Absätze von Erzeugnissen, die Bestandsveränderungen an Erzeugnissen sowie die selbsterstellten Anlagen abzüglich der Vorleistungen (insbesondere fremdbezogene Roh-, Hilfs- und Betriebsstoffe, vergebene Lohnarbeit)<sup>145</sup>.

In der Literatur werden verschiedene Größen zur Messung des Arbeitsinputs diskutiert: Zahl der Beschäftigten, Arbeitsstunden sowie Arbeitsvergütung (Löhne und Gehälter). Die Zahl der Beschäftigten ist die am einfachsten zu ermittelnde Größe. Grundlage der Simulation ist der Beschäftigtenstand am 25. Mai 1987 (Zeitpunkt der Arbeitsstättenzählung). Bei der Verwendung dieser rein quantitativen Maßzahl bleiben jedoch zeitliche Aspekte wie Teilzeitarbeit, Kurzarbeit, Überstunden usw. außer Betracht. Wird der Arbeitseinsatz durch die Arbeitszeit gemessen, dann wird der zeitlichen Dimension zwar Rechnung getragen, die Qualität der Arbeitsleistung bleibt jedoch auch bei diesem Meßkonzept außen vor. Um möglichen Qualitätsunterschieden der Arbeitsleistung Rechnung zu tragen, wird in der Literatur die Arbeitsvergütung als Indikator für unterschiedliche Arbeitsqualität vorgeschlagen (vgl. z.B. Weber 1983, S. 92). Dies impliziert jedoch, daß in den Entlohnungssystemen Qualitätsdifferenzen entsprechend differenziert berücksichtigt werden.

Mittels des vorliegenden Datensatzes kann die zeitliche und qualitative Dimension des Arbeitseinsatzes auf unterschiedliche Weise berücksichtigt werden. Obwohl der Anteil der Teilzeitbeschäftigten nicht explizit im Datensatz vorhanden ist, kann als Proxi für diesen Anteil der Frauenanteil an den Beschäftigten herangezogen werden, da Frauenanteil und Teilzeitanteil stark voneinander abhängig sind. Auf Ebene des Zweistellers (37 Untergruppen) der Arbeitsstättenzählung

---

<sup>144</sup> Da den folgenden Schätzungen eine synthetische Stichprobe zugrundeliegt, werden nicht nur Analysen zum unmittelbaren Testen der von Frick formulierten Hypothesen durchgeführt, sondern anhand von Faktornachfragefunktionen für die Zahl der Beschäftigten, den Frauenanteil, den Facharbeiteranteil sowie die Höhe des Kapitaleinsatzes wird der Datensatz auf Plausibilität überprüft. Dabei wird, wenn möglich, auf die Ergebnisse anderer Studien Bezug genommen.

<sup>145</sup> Vgl. z.B. Weber 1983, S. 6f.; Pietsch 1984, S. 86.

korrelieren die beiden Quoten in der Gesamtwirtschaft mit 0,65. Der Korrelationskoeffizient für das Verarbeitende Gewerbe beträgt sogar 0,92.

Zur Berücksichtigung der qualitativen Dimension kann zum einen die durchschnittliche Lohnhöhe und zum anderen der Anteil der Facharbeiter und Auszubildenden verwendet werden. Aufgrund humankapitaltheoretischer Konzepte müßte ein positiver Einfluß des Facharbeiteranteils und des Auszubildendenanteils auf die Bruttowertschöpfung pro Beschäftigtem zu beobachten sein. Wenn ein höherer Durchschnittslohn als Proxi für qualitativ bessere Arbeitsleistung angesehen werden kann, dann ist auch hier ein positiver Einfluß auf die Bruttowertschöpfung zu erwarten<sup>146</sup>.

Der Kapitaleinsatz wird durch die im Datensatz vorhandene Größe "Bruttoanlagevermögen" abgebildet, die zwar nur ein recht grobes Maß für die Kapitalintensität ist, weil Bestandswerte dieser Art implizit unterstellen, daß zwei Maschinen mit gleichen Anschaffungs- oder Herstellungskosten als äquivalent in ihrem Beitrag zum Output angesehen werden. Die verwendete Größe berücksichtigt zudem nicht das Alter des Anlagevermögens, d.h. Prozeßinnovationen werden nicht abgebildet. Da aber das Bruttoanlagevermögen pro Beschäftigtem in die Schätzung eingeht und für die Unternehmen die gleichen handelsrechtlichen Abschreibungsvorschriften gelten, kann daher von einem unterschiedlichen Niveau dieser Bestandsgröße für zwei Unternehmen mit sonst annähernd identischen Merkmalen auch auf die Innovationstätigkeit und -bereitschaft geschlossen werden.

In einem ersten Schritt werden nachfolgend die Ergebnisse der Schätzung einer Cobb-Douglas-Produktionsfunktion für den Zwei-Faktoren-Fall vorgestellt. Die Ergebnisse werden für die mit dem Startwert Null des Zufallszahlengenerators erzeugte synthetische Stichprobe dargestellt. Sofern sich große Abweichungen bei anderen Startwerten ergeben, wird in den Erläuterungen darauf hingewiesen. Die Ergebnisse für die anderen Startwerte sind den Tabellen 6-13 im Anhang zu entnehmen. Tabelle 28 zeigt die Mittelwerte und Standardabweichungen der in die erste Schätzung einbezogenen Größen. In Tabelle 29 ist das Ergebnis der Schätzung der Cobb-Douglas-Produktionsfunktion aufgeführt. Die durchgeführte Schätzung ergibt eine Schätzgleichung, die hochsignifikant ist, d.h. die Nullhypothese, daß kein Zusammenhang zwischen der abhängigen und den unabhängigen Variablen besteht, muß abgelehnt werden ( $\alpha < 0,01$ ). Die beiden bisher in die Regressionsgleichung aufgenommenen Variablen erklären bereits 55% der Gesamtstreuung der abhängigen Variablen. Dabei ist der Einfluß der Variablen "Anlagevermögen" mehr als dreimal so groß wie der Einfluß der Variablen "Beschäftigte". Diese Ergebnis stimmt mit den Ergebnissen von Unger (1986,

---

<sup>146</sup> Ein positiver Einfluß der Lohnhöhe auf die Bruttowertschöpfung kann auch als Hinweis auf die Gültigkeit effizienzlohntheoretischer Modelle interpretiert werden.

S. 87ff) überein, der für das Verarbeitende Gewerbe (auf Basis von Branchendaten) einen geringen Einfluß des Arbeitsinputs auf die Arbeitsproduktivität nachweist. Die Ergebnisse für die synthetischen Stichproben mit anderen Startwerten des Zufallszahlengenerators unterscheiden sich nur geringfügig in der Höhe der Koeffizienten und der Varianzaufklärung (vgl. Tabelle 7 im Anhang).

Um zum einen die qualitativen und zeitlichen Dimensionen des Arbeitsinputs und zum anderen den Einfluß der Schwerbehindertenquote auf die Bruttowertschöpfung zu analysieren, werden in einer zweiten Regressionsschätzung die Variablen Frauenanteil (Proxi für Anteil der Teilzeitbeschäftigten), Facharbeiter- und Auszubildendenanteil (Proxi für Humankapital, d.h. Qualität des Arbeitsinputs), Schwerbehindertenanteil (zum direkten Test der von Frick geführten Argumentation), Ausländeranteil sowie der Durchschnittslohn (als aus Effizienzlohnmodellen hergeleiteter Variablen oder als weiterer Indikator für die Qualität des Arbeitsinputs) mit aufgenommen.

Tabelle 28: Mittelwerte und Standardabweichungen der Variablen "Bruttowertschöpfung", "Bruttoanlagevermögen" und "Beschäftigte" sowie der logarithmierten Werte

Variable	Mittelwert	Standardabweichung
Bruttowertschöpfung	51.335,44	15.218,06
Bruttowertschöpfung (logarithmiert)	10,81	0,28
Anlagevermögen	103.651,90	46.577,98
Anlagevermögen (logarithmiert)	11,47	0,36
Beschäftigte	170,78	1.316,04
Beschäftigte (logarithmiert)	4,02	1,01

Tabelle 29: Cobb-Douglas-Produktionsfunktion für zwei Inputfaktoren

Variable	B	SE B	Beta	T
Beschäftigte <sup>1</sup>	0,0555	0,0029	0,2000	19,414**
Anlagevermögen <sup>1</sup>	0,5315	0,0079	0,6899	66,975**
Konstante	4,4850	0,0903		49,649**
R <sup>2</sup> *100	55,166			
F-Wert	2.643,658**			

\*\*: $\alpha < 0,01$

<sup>1</sup> logarithmierte Werte

Aufgrund der theoretischen Überlegungen in Abschnitt IV.4.4.2.1 müßte der Einfluß der Variablen "Frauenanteil" auf die Bruttowertschöpfung pro Beschäf-

tigtem negativ sein, da diese Variable als Proxi für den Anteil an Teilzeitbeschäftigten verwendet wird. Da die durchschnittliche Wertschöpfung pro Beschäftigtem durch Division der Bruttowertschöpfung in einer Branche und der Beschäftigtenzahl in der jeweiligen Branche ermittelt wird, ist ein negativer Zusammenhang zwischen durchschnittlicher Bruttowertschöpfung und dem Frauenanteil (als Proxi für Teilzeitbeschäftigung) aus rein rechnerischen Gründen zu erwarten, da die Teilzeitbeschäftigten bei der Berechnung des Durchschnittswerts mit gleicher Gewichtung wie Vollzeitbeschäftigte berücksichtigt werden. Aus einem negativen Einfluß dieser Größe kann daher nicht auf einen Produktivitätsverlust im Zusammenhang mit einem hohen Frauen- und/oder Teilzeitanteil geschlossen werden.

Für die Variablen "Facharbeiteranteil" und "Auszubildendenanteil" ist ein positiver Einfluß auf die Bruttowertschöpfung pro Beschäftigtem zu erwarten, da mit diesen Variablen die qualitative Dimension des Arbeitsinputs gemessen wird. Je höher der Facharbeiteranteil (Auszubildendenanteil<sup>147</sup>), desto größer ist die Arbeitsproduktivität und damit die Bruttowertschöpfung pro Beschäftigtem. Auszubildende stehen dem Unternehmen potentiell zu einem späteren Zeitpunkt als Facharbeiter zur Verfügung. Unternehmen mit einem hohen Auszubildendenanteil sehen in der "Eigenproduktion" von Facharbeitern offensichtlich die gegenüber dem "Fremdbezug" kostengünstigere Alternative zum kontinuierlichen Aufbau von betriebsspezifischem Humankapital.

Ausgehend von den theoretischen Überlegungen bei Frick ist für die Variable "Schwerbehindertenanteil" zumindest ein nicht signifikant negativer oder sogar signifikant positiver Einfluß auf die Bruttowertschöpfung zu erwarten<sup>148</sup>, da Frick im Unterschied zu Kohli/Wolf (1987), die den Anteil älterer Arbeitnehmer als ein Maß für den Bestand an betrieblichen Organisationskapital bezeichnen, die Schwerbehindertenquote als die geeignetere Operationalisierung für Organisationskapital ansieht (vgl. Frick 1992, S. 230). "Obgleich die beiden Variablen hoch korreliert sind, ist die Verwendung der Schwerbehindertenquote aus verschiedenen Gründen sinnvoller: Während es offensichtlich keinen Zusammenhang zwischen der individuellen Arbeitsleistung und dem Lebensalter gibt (vgl.

---

147 Wenn der Auszubildendenanteil zudem noch als Indikator für Organisationskapital akzeptiert wird, dann ist - die Tragfähigkeit des von Frick entwickelten theoretischen Konzepts vorausgesetzt - auch aufgrund dieses Wirkungszusammenhangs ein positiver Einfluß des Schwerbehindertenanteils auf die Bruttowertschöpfung zu erwarten.

148 Wenn das theoretische Konzept von Frick unternehmerische Entscheidungsprozesse korrekt abbildet, dann sollte der Schwerbehindertenanteil einen positiven Einfluß auf die Bruttowertschöpfung haben, wenn der Einfluß anderer wichtiger Determinanten eliminiert wird. Zumindest sollte in Unternehmen mit hohem Schwerbehindertenanteil keine geringere Bruttowertschöpfung zu beobachten sein als in Unternehmen mit niedrigem Schwerbehindertenanteil oder ohne Schwerbehinderte.

Naschold/Wagner/Rosenow 1989, S. 89), sind die individuelle Arbeitsproduktivität und das Vorliegen einer Behinderung negativ korreliert" (Frick 1992, S. 230). Von "Organisationskapital" kann dann gesprochen werden, wenn spezifische institutionelle Arrangements, beispielsweise in Form eines internen Arbeitsmarktes, eine "Überkompensation" der unterdurchschnittlichen Produktivität eines Teils der Beschäftigten durch eine demgegenüber größere Produktivität der übrigen Arbeitnehmer bewirken (vgl. ebenda, S. 230f.). Wolf/Kohli (1988, S. 193) sehen demzufolge in schwerbehinderten Arbeitnehmern durchaus eine "betriebliche Ressource", die wesentlich zur Internalisierung des Produktionszieles und zur Produktivitätssteigerung beitragen kann.

Der Ausländeranteil kann als zusätzliche Variable zur Messung der qualitativen Dimension des Arbeitsinputs aufgefaßt werden, da Ausländer überwiegend auf Arbeitsplätzen eingesetzt werden, die keinen formalen Berufsausbildungsabschluß erfordern. In einer weiter unten dargestellten Schätzung der Bestimmungsgründe des Durchschnittslohns pro Beschäftigten wird gezeigt, daß der Ausländeranteil einen leicht negativen Einfluß auf die Lohnhöhe hat. Es kann daher erwartet werden, daß mit steigendem Ausländeranteil die Bruttowertschöpfung pro Beschäftigtem sinkt.

Der Zusammenhang zwischen Durchschnittslohn pro Beschäftigtem und Bruttowertschöpfung pro Beschäftigtem müßte sowohl aus effizienzlohntheoretischen Überlegungen als auch aus der Verwendung dieser Größe als Proxi für die Qualität des Arbeitsinputs positiv sein. Durch die Berücksichtigung des Frauenanteils (Proxi für Teilzeitanteil) in der Schätzgleichung werden Einflüsse, die sich aus einem hohen Teilzeitbeschäftigtenanteil ergeben, eliminiert.

Die Berücksichtigung der zusätzlichen Variablen zur Charakterisierung des Arbeitsinputs läßt eine Erhöhung des Bestimmtheitsmaßes erwarten, sofern die Variablen zusätzliche Einflußfaktoren zur Erklärung der Bruttowertschöpfung darstellen. Zudem müßte der Erklärungsbeitrag der Variablen "Beschäftigtenzahl" abnehmen, da die zusätzlich in die Schätzung mit einbezogenen Größen nicht unabhängig von der Beschäftigtenzahl sind. Dabei ist die Korrelation der unabhängigen Variablen jedoch nicht so groß (die Toleranz-Werte<sup>149</sup> sind alle größer

---

<sup>149</sup> Der Toleranz-Wert ist ein Maß zur Überprüfung von Multikollinearität. Für jede unabhängige Variable wird das Bestimmtheitsmaß ermittelt, daß sich bei Regression einer unabhängigen Variable auf die übrigen unabhängigen Variable ergeben würde. Ist dieses Bestimmtheitsmaß gleich eins, dann bedeutet dies, daß sich eine unabhängige Variable durch eine Linearkombination der anderen Variablen erzeugen läßt. Folglich enthält die unabhängige Variable keine zusätzliche Information und kann somit nicht zur Erklärung der abhängigen Variable beitragen. Den Toleranz-Wert für die einzelnen Schätzkoeffizienten erhält man durch Subtraktion dieses Bestimmtheitsmaßes von eins. Toleranzwerte, die nahe an eins liegen, bedeuten keine Multikollinearität (vgl. Backhaus et. al. 1989, S. 35f.).

als 0,7; der Maximalwert liegt bei 1), daß aufgrund der Korrelationen verzerrte Schätzwerte resultieren können.

Tabelle 30: Mittelwerte und Standardabweichungen der Variablen "Frauenanteil", "Facharbeiteranteil", "Auszubildendenanteil", "Schwerbehindertenanteil", "Ausländeranteil" und "durchschnittliche Lohnhöhe pro Beschäftigtem"

Variable	Mittelwert	Standardabweichung
Frauenanteil	31,381	24,814
Facharbeiteranteil	30,190	19,092
Auszubildendenanteil	7,036	5,313
Schwerbehindertenanteil	4,370	4,183
Ausländeranteil	9,907	6,562
Lohn	34.956,915	6.623,841
Lohn (logarithmiert)	10,442	0,202

Tabelle 31: Produktionsfunktion mit differenziertem Arbeitsinput

Variable	B	SE B	Beta	T
Lohn <sup>1</sup>	0,1911	0,01657	0,138	11,539**
Anlagevermögen <sup>1</sup>	0,5279	0,00777	0,685	67,932**
Facharbeiteranteil	-0,0011	0,00015	-0,080	-8,000**
Schwerbehindertenanteil	-0,0015	0,00067	-0,023	-2,316*
Ausländeranteil	-0,0016	0,00043	-0,038	-3,766**
Auszubildendenanteil	0,0067	0,00053	0,127	12,590**
Frauenanteil	-0,0006	0,00012	-0,052	-4,812**
Beschäftigte	0,0418	0,00306	0,151	13,694**
Konstante	2,6145	0,19156		13,649**
R <sup>2</sup> *100	58,964			
F-Wert	770,695**			

\*\*: $\alpha < 0,01$ ; \*: $\alpha < 0,05$

<sup>1</sup> logarithmierte Werte

Die in der Tabelle 31 gezeigten Schätzkoeffizienten sind alle statistisch signifikant und weisen bis auf "Facharbeiteranteil" und "Schwerbehindertenanteil" das prognostizierte Vorzeichen auf. Der Erklärungsbeitrag des Bruttoanlagevermögens bleibt nahezu unverändert, wohingegen der Erklärungsbeitrag der Beschäftigtenzahl geringer wird. Die Zunahme der Varianzaufklärung auf ca. 59% fällt relativ gering aus, was darauf hindeutet, daß von den zusätzlich in die Regressions-schätzung aufgenommenen Variablen ein Teil der Varianz erklärt wird, die in

der Schätzung des Zwei-Faktoren-Modells von der aggregierten Größe "Beschäftigtenzahl" erklärt wurde. Der immer noch sehr hohe Erklärungsbeitrag der Variablen "Beschäftigtenzahl" legt die Vermutung nahe, daß die zusätzlich in die Regression aufgenommenen Variablen den Arbeitsinput nicht in seiner ganzen Komplexität erklären bzw. der rein quantitative Einfluß dominiert, d.h. die qualitative Dimension des Arbeitsinputs von untergeordneter Bedeutung ist. Für diese Interpretation spricht auch der nach wie vor unverändert hohe Erklärungsbeitrag der Variablen "Bruttoanlagevermögen".

Die in Tabelle 31 aufgeführten Ergebnisse sind auf Basis der synthetischen Stichprobe mit dem Startwert Null des Zufallszahlengenerators berechnet worden. Für andere Startwerte des Zufallszahlengenerators sind die Ergebnisse bis auf die Variable "Frauenanteil", der in einer von sechs konstruierten Stichproben insignifikant ist, sowohl in der Richtung des Zusammenhangs als auch in der quantitativen Ausprägung der Regressionskoeffizienten sehr ähnlich (vgl. Tabelle 8 im Anhang). Der Schwerbehindertenanteil hat in drei der sechs Stichproben sogar einen negativen Einfluß auf die Bruttowertschöpfung auf dem 99%-Niveau ( $\alpha < 0,01$ ).

Der signifikant negative Einfluß des Schwerbehindertenanteils auf die Bruttowertschöpfung pro Beschäftigtem läßt darauf schließen, daß der von Frick und anderen Autoren postulierte Zusammenhang zwischen Organisationskapital, für das der Schwerbehindertenanteil als Indikator gilt, und Produktivität für das Verarbeitende Gewerbe nicht direkt nachgewiesen werden kann. Frick (1992, S. 244) kommt auf Basis von Berechnungen mit Individualdaten jedoch zu dem Ergebnis, "daß Schwerbehinderte überwiegend in Branchen und Betrieben weiterbeschäftigt (Hervorhebung i. Orig.) werden, in denen die aufeinander abgestimmten Rekrutierungs-, Qualifizierungs- und Entlohnungspraktiken interner Arbeitsmärkte eine Stabilisierung individueller Erwerbskarrieren und Berufsverläufe und damit gleichzeitig eine Reduktion von Arbeits- und Organisationskosten bewirken". Diese Hypothese wird anhand der Schätzung einer Faktornachfragefunktion (vgl. Abschnitt 4.4.3) überprüft, indem Variable, die einen internen Arbeitsmarkt charakterisieren, mit in die Schätzgleichung aufgenommen werden.

Aus dem signifikant negativen Regressionskoeffizienten für den Schwerbehindertenanteil sollte jedoch nicht auf einen mit einem hohen Schwerbehindertenanteil einhergehenden generellen Produktivitätsnachteil geschlossen werden. Da es sich bei den zur Schätzung verwendeten Daten um eine synthetische Stichprobe handelt, der als Konstruktionsprinzip das "Prinzip der maximalen Entropie" zugrundeliegt, kann der "wahre" Zusammenhang unter Umständen sogar positiv bzw. stärker negativ sein. Die Schätzergebnisse zeigen jedoch, daß eine Verwendung von Unternehmensdaten mit einer direkten Gegenüberstellung von Schwerbehindertenanteil (und anderen Variablen) mit Produktivitätsmaßen zu Ergebnissen führen, die nicht unmittelbar mit den vorgestellten theoretischen Konzepten im

Einklang stehen. Möglicherweise würden sich die Ergebnisse jedoch verändern, wenn weitere Variable wie Fluktuationsrate und Betriebszugehörigkeitsdauer mit in die Regressionsschätzung aufgenommen werden.

Sehr kontraintuitiv ist der hochsignifikant negative Einfluß des Facharbeiteranteils auf die Bruttowertschöpfung pro Beschäftigtem unter anderem deshalb, weil der Einfluß einiger anderer als relevant anzusehender Faktoren (Bruttoanlagevermögen, Lohnhöhe, Frauenanteil) in der Schätzgleichung berücksichtigt wird. In vielen anderen Studien werden existierende Lohndifferentiale durch Berücksichtigung des Qualifikationsniveaus der Beschäftigten in den Schätzgleichungen teilweise erklärt<sup>150</sup> (vgl. unten die Regressionsergebnisse zu den Determinanten der Lohnhöhe). Wenn Lohnhöhe und Arbeitsleistung in Zusammenhang stehen (wofür auch der unten aufgeführte Einfluß der Bruttowertschöpfung auf die Lohnhöhe spricht), dann müßte der Einfluß des Facharbeiteranteils auf die Bruttowertschöpfung positiv sein, sofern der Facharbeiteranteil in einem positiven Zusammenhang mit der Arbeitsleistung steht. Wenn der Facharbeiterstatus mit einem höheren allgemeinen und/oder betriebsspezifischen Humankapital korreliert, dann ist ein positiver Zusammenhang zwischen Facharbeiteranteil und Arbeitsleistung zu erwarten. Auch wenn aufgrund bisheriger empirischer Ergebnisse (vor allem auf Basis von Individualdaten) von einem positiven Einfluß des Qualifikationsniveaus auf die individuelle Lohnhöhe gesprochen werden muß, kann daraus jedoch nicht gefolgert werden, daß Unternehmen, die einen höheren Facharbeiteranteil haben, eine höhere Bruttowertschöpfung pro Beschäftigtem aufweisen.<sup>151</sup>

Weitergehenden Aufschluß über die Determinanten des Facharbeiter- und Schwerbehindertenanteils erhält man durch eine Schätzung von Faktornachfragefunktionen. Zunächst soll anhand der aus dem Zwei-Faktoren-Modell abgeleiteten Faktornachfragefunktion (vgl. Gleichung 42) eine Überprüfung der Plausibilität des Datensatzes erfolgen, bevor dann die Nachfragefunktionen für den Schwerbehinderten- und Facharbeiteranteil (und zum Vergleich auch die Funktionen für den Frauenanteil und das Bruttoanlagevermögen) geschätzt und einer detaillierteren Analyse unterzogen werden. In diesem Modell ist die Arbeitsnachfrage nur abhängig vom Outputniveau  $Y$  und vom realen Faktorpreis. Da das

---

<sup>150</sup> Vgl. z.B.: Gerlach/Schmidt 1989, S. 359; Frick 1992, S. 181; Kruse 1991, S.235; Groshen 1991, S. 879; Brown/Medoff 1989, S. 1036ff; Evans/Leighton 1988, S. 310f.

<sup>151</sup> Um dem Einwand zu begegnen, daß der geschätzte Zusammenhang durch die Disaggregation bei der Erzeugung der synthetischen Stichprobe "hineinsimuliert" wurde, soll darauf hingewiesen werden, daß der Korrelationskoeffizient zwischen Facharbeiteranteil und Bruttowertschöpfung auf Ebene der neun Untergruppen aus der Arbeitsstättenzählung für das Verarbeitende Gewerbe  $-0,31$  beträgt, d.h. auch auf Basis der Aggregatdaten zeigt sich ein negativer univariater Zusammenhang, der auch in einem multivariaten Schätzmodell auf Basis von synthetischen Unternehmensdaten bestehen bleibt.

Preisniveau für alle Unternehmen gleich ist, kann der Nominallohn als Faktorpreis verwendet werden. Als abhängige Variable (Arbeitsnachfrage) wird der Logarithmus der Beschäftigtenzahl verwendet. Die geschätzten Koeffizienten sollten Werte aufweisen, die nahe bei eins (für  $\ln Y$ ) und nahe bei -1 ( $\ln w$ ) liegen.

Tabelle 32: Arbeitsnachfragefunktion für das Zwei-Faktoren-Modell

Variable	B	SE B	Beta	T
Wertschöpfung	0,7781	0,0498	0,2160	15,615**
Lohn	1,8125	0,0689	0,3640	26,318**
Konstante	-23,3202	0,8073		-28,888**
$R^2 \cdot 100$	21,084			
F-Wert	574,022**			

\*\*: $\alpha < 0,01$

Tabelle 32 ist zu entnehmen, daß der Regressionskoeffizient der Variable "Wertschöpfung" das erwartete Vorzeichen aufweist. Der positive Einfluß der Lohnhöhe auf die Nachfrage nach Arbeit steht zu dem theoretisch abgeleiteten negativen Einfluß dieser Größe auf die Arbeitsnachfrage im Widerspruch, da die Lohnhöhe als Faktorpreis des Produktionsfaktors Arbeit aufgefaßt wird. Möglicherweise ist die Größe "Durchschnittslohn pro Beschäftigtem" eine nicht unmittelbar geeignete Operationalisierung für die tatsächlichen Faktorkosten. Der positive Einfluß der Lohnhöhe deutet jedoch auch auf ein bisher in der ökonomischen Theorie und Empirie ungelöstes Phänomen hin: Offenbar zahlen große Unternehmen höhere Löhne als kleine und mittlere Unternehmen. In einer Vielzahl von empirischen Analysen auf Basis von Individual-, Unternehmens- und Branchendaten konnte selbst bei Berücksichtigung einer Vielzahl potentieller Einflußfaktoren das Betriebsgrößenlohndifferential nicht annähernd vollständig erklärt werden<sup>152</sup>.

Zur genaueren Analyse der Bestimmungsgründe der Lohnhöhe bietet sich die Möglichkeit, eine Regressionsschätzung mit der Lohnhöhe als abhängiger Variable durchzuführen. Die nachfolgenden Überlegungen und Regressionsanalysen stellen weder fundamentale lohntheoretische Überlegungen dar, noch soll ein detaillierter Überblick über den Stand der lohntheoretischen Diskussion gegeben werden. Die Schätzung des Modells dient einerseits zur Validierung der in der Stichprobe abgebildeten Zusammenhänge, und andererseits soll aufgrund der spezifizierten Schätzgleichung eine differenziertere Darstellung des Faktorpreises

<sup>152</sup> Einen guten Überblick über die theoretischen Erklärungsmodelle zu diesem Phänomen geben Gerlach/Schmidt 1989, S. 355ff. Zur empirischen Analyse vgl. z.B.: ebenda, S. 358ff. Siehe auch: Brown, C./Medoff, J. 1989, Groshen, E.L. 1991.

für den Produktionsfaktor Arbeit ermöglicht werden. Die gefundenen Zusammenhänge werden dann genutzt, um eine differenziertere Arbeitsnachfragefunktion zu schätzen.

Folgende unabhängige Variable werden in die Regressionsschätzung mit aufgenommen: Wertschöpfung, Frauenanteil, Behindertenanteil, Facharbeiteranteil, Auszubildendenanteil, Ausländeranteil, Beschäftigte und Schwerbehindertenanteil. Die Wertschöpfung mißt dabei den Einfluß der Produktivität auf die Höhe der Entlohnung, und es müßte daher ein positiver Zusammenhang zu beobachten sein. Durch die Berücksichtigung des Frauenanteils wird der Einfluß des Teilzeitbeschäftigtenanteils kontrolliert, da mit höherem Teilzeitanteil der Durchschnittslohn sinken müßte<sup>153</sup>. Sowohl die Variable Ausländeranteil als auch die Variable Auszubildendenanteil müßten einen negativen Einfluß auf die Lohnhöhe haben. Auszubildende erhalten eine Ausbildungsvergütung, die in der Regel weit unter der durchschnittlichen Lohnhöhe pro Beschäftigtem liegen wird, so daß sich rein rechnerisch schon ein negativer Einfluß zeigen sollte.

Der Facharbeiteranteil wird als Proxivariable für die Humankapitalausstattung der Beschäftigten mit in die Regressionsschätzung aufgenommen, d.h. mit steigendem Facharbeiteranteil sollte auch der Durchschnittslohn steigen. In vielen empirischen Studien zu den Determinanten der Individualeinkommen zeigt sich auf Basis verschiedener Datensätze in unterschiedlichen Ländern ein signifikant positiver Einfluß der Humankapitalausstattung auf das individuelle Erwerbseinkommen. Unter der Voraussetzung, daß der Facharbeiterstatus als Proxi für die Humankapitalausstattung eine valide Operationalisierung der Humankapitalausstattung ist, sollte auch auf Unternehmensebene ein positiver Einfluß auf die durchschnittliche Lohnhöhe pro Beschäftigtem zu beobachten sein.

Da in vielen empirischen Studien<sup>154</sup> ein großer Einfluß der Betriebsgröße auf den Lohn nachgewiesen werden konnte, der selbst bei Berücksichtigung einer Vielzahl möglicher anderer Einflußfaktoren bestehen bleibt, wird die Variable Beschäftigtenzahl mit in die Regressionsschätzung aufgenommen. Wenn die Schätzergebnisse, selbst bei Berücksichtigung anderer wesentlicher Einflußgrößen, wie z. B. der Wertschöpfung, das in der Literatur vielfach nachgewiesene,

---

<sup>153</sup> Wenn auch der Frauenanteil als Proxi für den Anteil der Teilzeitbeschäftigten verwendet wird, kann nicht davon ausgegangen werden, daß ein potentiell negativer Einfluß des Frauenanteils ausschließlich auf den größeren Anteil an Teilzeitbeschäftigten zurückzuführen ist. Es liegt vielmehr die Vermutung nahe, daß Frauen nach wie vor hinsichtlich ihres Einkommens diskriminiert werden. Sowohl die von Frick auf Basis des Sozioökonomischen Panels berechneten Segregationsindizes als auch die auf Basis der synthetischen Stichprobe berechneten Segregationsindizes deuten auf eine Beschäftigungsdiskriminierung von Frauen hin, die möglicherweise mit einer Einkommensdiskriminierung einhergeht.

<sup>154</sup> Vgl. z. B.: Brown/Medoff 1989, Groshen 1991, Kruse 1992, Gerlach/Schmidt 1989.

betriebsgrößenabhängige Lohndifferential zeigen, kann dies als Indiz für die Qualität der Abbildung der Grundgesamtheit durch die synthetische Stichprobe angesehen werden.

In die Schätzung wird zudem noch ein Maß für die Kapitalproduktivität mit einbezogen. Es ist zu erwarten, daß mit zunehmender Kapitalproduktivität der Lohn steigt. Die Kapitalproduktivität wird durch den Quotienten aus durchschnittlicher Wertschöpfung und durchschnittlichem Anlagevermögen operationalisiert.

Der Schwerbehindertenanteil sollte aufgrund der Ergebnisse von Frick (1992, S. 199f.) einen positiven Einfluß auf die durchschnittliche Lohnhöhe haben, da Schwerbehinderte ein höheres Durchschnittsalter aufweisen als Nichtbehinderte und Schwerbehinderte daher eine längere Betriebszugehörigkeitsdauer besitzen als Nichtbehinderte. Sadowski/Frick (1992, S. 56f.) zeigen zudem, daß der Anteil der intern rekrutierten Schwerbehinderten (d.h. bereits beschäftigte, gesundheitlich beeinträchtigte Arbeitnehmer werden nach der Anerkennung als Schwerbehinderte weiterbeschäftigt) ca. 75% beträgt. Da die Weiterbeschäftigung im gleichen Unternehmen wohl kaum zu einem niedrigeren Lohn führt, ist von einem positiven Einfluß der Schwerbehindertenquote auf die Lohnhöhe auszugehen.

Frick (1992, S. 243f.) kommt zu dem Ergebnis, daß der Umfang der Schwerbehindertenbeschäftigung im wesentlichen von der Bedeutung interner Arbeitsmärkte abhängt. Ein wesentliches Charakteristikum interner Arbeitsmärkte sind an der Betriebszugehörigkeitsdauer orientierte Entlohnungssysteme, d.h. die Lohnhöhe nimmt mit der Betriebszugehörigkeitsdauer zu. Wegen der im Vergleich zu nichtbehinderten Arbeitnehmern höheren Betriebszugehörigkeitsdauer und deren positivem Einfluß auf die Lohnhöhe sollte mit steigendem Schwerbehindertenanteil ein höherer Durchschnittslohn pro Arbeitnehmer zu beobachten sein.

Tabelle 33 zeigt die Ergebnisse der Regression zu den Determinanten der durchschnittlichen Lohnhöhe im Unternehmen. Bis auf die Variable Ausländeranteil weisen alle in die Schätzung mit aufgenommenen Variablen einen signifikanten Einfluß auf. Die Variablen mit signifikantem Einfluß besitzen die postulierte Wirkungsrichtung. Trotz Berücksichtigung anderer als wesentlich erachteter Einflußfaktoren (insbesondere der Wertschöpfung, der Kapitalproduktivität und des Frauenanteils) kann das Betriebsgrößenlohndifferential nicht annähernd erklärt werden. Der hochsignifikant positive Einfluß der Variablen "Beschäftigte" sowie der hohe standardisierte Regressionskoeffizient, der den relativen Erklärungsbeitrag der Variablen mißt, bestätigen dieses Ergebnis. Neben der Betriebsgröße weist die Variable "Frauenanteil" den größten relativen Erklärungsbeitrag auf. Einerseits bedeutet dies, daß in der synthetischen Stichprobe der Zusammenhang zwischen Teilzeitbeschäftigung und durchschnittlicher Lohnhöhe korrekt abgebildet wird (rechnerischer Effekt), und andererseits liegt die Vermutung nahe, daß Frauen hinsichtlich ihres Einkommens möglicherweise diskriminiert wer-

den. Die Diskriminierungshypothese wird durch die Ergebnisse der geschätzten Produktionsfunktion gestützt. Obwohl der Koeffizient der Variable "Frauenanteil" in der Produktionsfunktion (vgl. Tabelle 31) ein negatives Vorzeichen aufweist und statistisch signifikant<sup>155</sup> ist, erweist sich der relative Erklärungsbeitrag dieser Variable für die durchschnittliche Bruttowertschöpfung als der niedrigste von allen in die Schätzung der Bruttowertschöpfung einbezogenen Variablen. Es kann daher davon ausgegangen werden, daß der negative Einfluß auf die Bruttowertschöpfung aus dem oben dargestellten rechnerischen Effekt resultiert (hoher Anteil an Teilzeitbeschäftigten).

Sowohl der aus der neoklassischen Theorie abgeleitete Zusammenhang zwischen Wertschöpfung und Entlohnung als auch der aus humankapitaltheoretischen Überlegungen postulierte Zusammenhang zwischen Humankapital und Entlohnung kann durch die Schätzung auf Basis der synthetischen Stichprobe belegt werden. Die Ergebnisse der Schätzungen decken sich in diesen Punkten mit den in der Literatur vielfach dargestellten Ergebnissen, d.h. die Zusammenhänge zwischen den einzelnen Größen in der synthetischen Stichprobe sind, obwohl das Konstruktionsprinzip der "maximalen Entropie" zugrundeliegt, nicht beliebig, sondern nähern sich den in der Grundgesamtheit vorhandenen Zusammenhängen an.

Die oben dargestellten Ergebnisse zu den Determinanten der Lohnhöhe können zu einer differenzierteren Schätzung der Arbeitsnachfrage herangezogen werden. Berücksichtigt man in der Schätzung der Arbeitsnachfragefunktion zusätzlich zu den in Tabelle 32 dargestellten Variablen "Lohn" und "Bruttowertschöpfung" die Variablen, die zur Erklärung der Lohnhöhe herangezogen werden, kann der neoklassisch postulierte Zusammenhang zwischen Lohnhöhe und Arbeitsnachfrage möglicherweise nachgewiesen werden. Der positive Einfluß der Lohnhöhe auf die Arbeitsnachfrage, der konträr zu der in Abschnitt IV.4.4.2.1 postulierten Wirkungsrichtung ist, liegt vielleicht darin begründet, daß die Lohnhöhe mit der Unternehmensgröße zunimmt. Es handelt sich also möglicherweise um einen durch das bislang theoretisch und empirisch ungeklärte Betriebsgrößenlohndifferential erzeugten Effekt, d.h. aus den Schätzergebnissen der Arbeitsnachfrage muß nicht auf ein nichtzutreffendes Modell oder eine den Zusammenhängen in der Grundgesamtheit nicht gerecht werdende synthetische Stichprobe geschlossen werden.

Aufgrund der Schätzergebnisse zu den Determinanten der Lohnhöhe ist im Falle einer Berücksichtigung der erklärenden Variablen bei der Schätzung der Arbeitsnachfragefunktion ein positiver Einfluß des Frauenanteils (stark negativer Einfluß

---

<sup>155</sup> Der Regressionskoeffizient des Frauenanteils in der Produktionsfunktion ist in zwei Stichproben (Startwerte eins und zwei des Zufallszahlengenerators) insignifikant, d.h. die Schätzergebnisse auf Basis der synthetischen Stichprobe lassen nicht eindeutig auf einen negativen Einfluß des Frauenanteils auf die Bruttowertschöpfung schließen.

auf die Lohnhöhe) zu erwarten: ein hoher Frauenanteil verringert den Preis für den Faktor Arbeit. Für die Richtung des Wirkungszusammenhangs zwischen Facharbeiteranteil und Arbeitsnachfrage ist zum einen wegen des höheren Lohns von Facharbeitern und zum anderen wegen Substitutionseffekten mit einem negativen Vorzeichen zu rechnen. Aufgrund der Schätzergebnisse zu den Determinanten des Durchschnittslohns (der Auszubildendenanteil verringert den Durchschnittslohn) müßte das Vorzeichen der Variablen "Auszubildendenanteil" positiv sein. Da aber der Status des Auszubildenden nur temporär ist und im Anschluß an eine Ausbildung Facharbeiterlöhne gezahlt werden, könnte der positive Einfluß kompensiert werden, d.h. über die Richtung des Wirkungszusammenhangs kann keine begründete Hypothese formuliert werden.

Tabelle 33: Determinanten des Durchschnittslohns pro Beschäftigten

Variable	B	SE B	Beta	T
Kapitalproduktivität	0,0020	0,00021	0,123	9,750**
Auszubildendenanteil	-0,0062	0,00048	-0,163	-12,989**
Schwerbehindertenanteil	0,0018	0,00060	0,037	2,960**
Ausländeranteil	0,0003	0,00039	0,010	0,775
Facharbeiteranteil	0,0009	0,00013	0,089	7,137**
Beschäftigte	0,0690	0,00262	0,344	26,361**
Frauenanteil	-0,0027	0,00010	-0,331	-26,217**
Bruttowertschöpfung	0,0702	0,00953	0,097	7,362**
Konstante	9,3886	0,10101		92,948**
$R^2 \cdot 100$	35,768			
F-Wert	298,682**			

\*\*. $\alpha < 0,01$

Die Ergebnisse der geschätzten Arbeitsnachfragefunktion sind in Tabelle 34 dargestellt. Gegenüber dem Zwei-Faktoren-Modell (vgl. Tabelle 32) erhöht sich die Varianzaufklärung nur geringfügig, da in die Gleichung Variable mit einbezogen werden, die auf die Lohnhöhe einen signifikanten Einfluß haben. Der positive Einfluß der Bruttowertschöpfung auf die Arbeitsnachfrage bleibt auch im erweiterten Modell mit fast identischen Regressionskoeffizienten und T-Wert erhalten. Auch der positive Einfluß der Lohnhöhe verändert sich nur unwesentlich. Die neu mit in die Schätzung einbezogenen Variablen haben bis auf den Auszubildendenanteil einen signifikanten Einfluß auf die Arbeitsnachfrage. Die Regressionskoeffizienten der Variablen "Frauenanteil" und "Facharbeiteranteil" weisen die aufgrund der geschätzten Gleichung für den Durchschnittslohn erwarteten Vorzeichen auf, d.h. je höher der Anteil der Facharbeiter im Unternehmen, desto niedriger die Arbeitsnachfrage, bzw. je höher der Frauenanteil, desto höher die

Arbeitsnachfrage. Diese Ergebnisse stehen im Einklang mit den in Abschnitt IV.4.4.2.1 abgeleiteten theoretischen Überlegungen in dem Sinne, daß die Arbeitsnachfrage steigt (sinkt), wenn der Anteil der Beschäftigten mit niedrigem (hohem) Durchschnittslohn zunimmt (abnimmt). Die Zusammenhänge in der synthetischen Stichprobe halten einem Vergleich mit Schätzungen auf Basis von "echten" Unternehmensdaten durchaus stand. Breitung (1992, S. 169ff) kommt auf Basis eines Paneldatensatzes von Unternehmen mit mehr als 20 Beschäftigten des Investitionsgütersektors in Niedersachsen ( $n = 920$ ) für den Einfluß der Bruttowertschöpfung zu ähnlichen Ergebnissen. Der Autor verwendet als Operationalisierung für die Bruttowertschöpfung den Bruttoproduktionswert aller im Unternehmen hergestellten Güter und als Operationalisierung für den Preis des Faktors Arbeit die Lohnsumme geteilt durch die Summe der geleisteten Arbeiterstunden. Der Koeffizient der Variablen "Bruttoproduktionswert" liegt in dieser Schätzung bei 0,609 (synthetische Stichprobe: 0,778).

Tabelle 34: Arbeitsnachfragefunktion mit differenzierten erklärenden Variablen

Variable	B	SE B	Beta	T
Lohn	2,0710	0,0754	0,416	27,462**
Ausländeranteil	0,0077	0,0021	0,050	3,718**
Schwerbehindertenanteil	0,0192	0,0032	0,080	5,938**
Facharbeiteranteil	-0,0021	0,0007	-0,041	-3,006**
Auszubildendenanteil	0,0015	0,0026	0,008	0,574
Bruttowertschöpfung	0,7633	0,0506	0,212	15,079**
Frauenanteil	0,0060	0,0006	0,148	10,149**
Konstante	-26,1537	0,8574		-30,500**
$R^2 \cdot 100$	23,975			
F-Wert	193,362**			

\*\*: $\alpha < 0,01$

Verwendet man den Bruttoproduktionswert auf Unternehmensebene aus der synthetischen Stichprobe zur Schätzung der Arbeitsnachfrage, dann erhöht sich die Varianzaufklärung auf 94,611 und der Regressionskoeffizient hat einen Wert von 0,875. Der Wert des Regressionskoeffizienten weist damit eine ähnliche Größenordnung auf wie der theoretisch prognostizierte Wert (1). Die Schätzung einer Arbeitsnachfragefunktion auf Basis der Bruttowertschöpfung pro Unternehmen erscheint jedoch als inadäquate Spezifikation, da die Zahl der Beschäftigten und der Bruttoproduktionswert zwangsläufig hoch miteinander korrelieren, sofern Unternehmen nicht dauerhaft in der Verlustzone produzieren. In einer Schätzung verwendete Durchschnittswerte eliminieren den rein quantitativen

Einfluß der Beschäftigtenzahl auf die Wertschöpfung im Unternehmen, da nur die durchschnittliche Produktivität betrachtet wird.

Der positive Einfluß des Schwerbehindertenanteils auf die Arbeitsnachfrage ist möglicherweise durch eine umgekehrte Kausalität zwischen Unternehmensgröße und Schwerbehindertenanteil zu erklären: große Unternehmen weisen einen größeren Anteil an schwerbehinderten Arbeitnehmern auf als kleine und mittelgroße Unternehmen (vgl. hierzu die Ergebnisse in Abschnitt IV.4.4.3, vgl. auch Sadowski/Frick 1992, S. 53).

Sowohl die Schätzung einer Arbeitsnachfragefunktion als auch die ergänzende Bestimmung der Determinanten der Lohnhöhe führen zu Ergebnissen, die einerseits mit den theoretisch postulierten Zusammenhängen im Einklang stehen und andererseits den auf Basis anderer Datensätze (echte Unternehmensdaten) ermittelten Ergebnissen nicht widersprechen. Es ergeben sich aus den bisherigen Ergebnissen zur Faktornachfrage keine Hinweise darauf, daß die in der synthetischen Stichprobe enthaltenen Strukturinformationen nicht den in der Grundgesamtheit der Unternehmen vorhandenen Zusammenhängen entsprechen. Abschließend erfolgt daher zur Bestimmung der Determinanten der Schwerbehindertenbeschäftigung die Schätzung einer Nachfragefunktion für Schwerbehinderte. Die Ergebnisse dieser Schätzung können mit den Ergebnissen von Frick (1992, S. 226ff) verglichen werden, so daß näher untersucht werden kann, ob Schätzungen auf Basis von Unternehmensdaten zu anderen Ergebnissen führen als Schätzungen auf Basis von Branchendaten.

#### **4.4.3 Bestimmungsgründe der betrieblichen Schwerbehindertenbeschäftigung**

Um die Bestimmungsgründe der Schwerbehindertenbeschäftigung zu ermitteln, wird in zwei Schritten vorgegangen. Da der Anteil der Unternehmen, die keinen einzigen Schwerbehinderten beschäftigen, in den synthetischen Stichproben bei ca. 25% liegt (vgl. Tabelle 25), wird zunächst eine logistische Regression mit der dichotomen Variable "Schwerbehinderte beschäftigt (ja/nein)" als abhängiger Variable durchgeführt. Als unabhängige Variable gehen in diese Schätzung alle Variablen der synthetischen Stichprobe ein, die einen Einfluß auf die betriebliche Schwerbehindertenquote haben könnten. In einem zweiten Schritt erfolgt dann die Schätzung einer Faktornachfragefunktion für den Schwerbehindertenanteil mit dem Schwerbehindertenanteil als abhängiger und den in der oben erwähnten logistischen Regression verwendeten Variablen als erklärenden Variablen. Zusätzlich werden die Residuen der logistischen Regression zusätzlich als erklärende Variable berücksichtigt, so daß der Einfluß der Unternehmen ohne beschäftigte Schwerbehinderte kontrolliert werden kann. Die Residualgröße aus der logistischen Regression beschreibt dabei die nicht durch die anderen unabhängigen Va-

riablen erklärte Wahrscheinlichkeit, Schwerbehinderte zu beschäftigen. Die Ergebnisse (vgl. Tabelle 35 und Tabelle 12 im Anhang) der Regressionsschätzung mit der Schwerbehindertenquote als abhängiger Variable sind sehr heterogen für die mit unterschiedlichen Startwerten ermittelten Stichproben. Einen erwartungsgemäß hohen und auch signifikanten Erklärungsbeitrag in allen Stichproben zeigt die Residualgröße aus der logistischen Regression. Gegeben die Wahrscheinlichkeit überhaupt einen Schwerbehinderten zu beschäftigen, haben nur die Variablen "Bruttowertschöpfung" und "Beschäftigtenzahl" in allen Stichproben einen signifikanten Einfluß auf die Schwerbehindertenquote. Der Einfluß der Bruttowertschöpfung auf die Schwerbehindertenquote ist negativ, während mit der Beschäftigtenzahl die Schwerbehindertenquote ansteigt. In diesem Zusammenhang ist die negative Korrelation zwischen Bruttowertschöpfung und Schwerbehindertenquote keinesfalls überraschend, da die Schätzung der Produktionsfunktion dies bereits in umgekehrter Kausalität zeigte. Frick (1992) ermittelt auf Basis der Branchendaten einen signifikant negativen Einfluß der Fluktuationsrate und einen signifikant positiven Einfluß der Stabilitätsquote auf die Schwerbehindertenquote<sup>156</sup>. Keine der anderen in eine direkte Regressionsschätzung zur Erklärung der Schwerbehindertenquote mit einbezogenen Variablen<sup>157</sup> zeigt auf Basis der Branchendaten einen signifikanten Einfluß. In einem Pfadmodell gelangt Frick zu dem Ergebnis, daß die Beschäftigtenzahl und das allgemeine Humankapital auf die Stabilitätsquote und damit auf die Schwerbehindertenquote einen positiven Einfluß aufweisen. Verstärkt wird der Einfluß der Betriebsgröße auf die Schwerbehindertenquote bei Frick durch deren negativen Einfluß auf die Fluktuationsrate, die wiederum negativ mit der Schwerbehindertenquote zusammenhängt. Zudem wirkt die Frauenquote auf die Stabilitätsrate und die Fluktuationsrate negativ, so daß der Einfluß auf die Schwerbehindertenquote ungewiß bleibt. Die Konzentrationsrate bewirkt über den negativen Einfluß auf die Arbeitslosenquote indirekt eine Erhöhung der Schwerbehindertenquote.

Die Ergebnisse auf Basis der synthetischen Stichprobe, die sich nur auf das Verarbeitende Gewerbe beziehen, stützen teilweise die Ergebnisse von Frick, widersprechen diesen aber auch in einigen wesentlichen Punkten, bzw. es kann ein Zusammenhang nicht nachgewiesen werden. Auf Basis der Branchendaten war der Zusammenhang zwischen Bruttowertschöpfung und Schwerbehindertenquote

---

<sup>156</sup> Die Fluktuationsrate wird operationalisiert als die Nettoersatzungsrate sozialversicherungspflichtiger Beschäftigungsverhältnisse von 1980-1984. Als Operationalisierung für die Stabilitätsquote verwendet Frick den Anteil der über Fünfzigjährigen an allen sozialversicherungspflichtig Beschäftigten (vgl. Frick 1992, S. 218f.).

<sup>157</sup> Folgende Variablen wurden von Frick in die Regressionsschätzung mit aufgenommen: Betriebsgröße, Konzentration, Frauenquote, Beschäftigungsstabilität, Fluktuationsrate, allgemeines Humankapital, Arbeitslosenquote, Beschäftigungsentwicklung und Pro-Kopf-Einkommen.

nicht testbar, so daß hier kein Vergleich möglich ist. Übereinstimmend kann in beiden Datensätzen der Zusammenhang zwischen Betriebsgröße und Schwerbehindertenquote gezeigt werden. Auch der nicht signifikante Einfluß der Frauenquote auf Basis der synthetischen Stichprobe widerspricht nicht den Resultaten von Frick. Die Konzentrationsrate jedoch hat auf Basis der synthetischen Stichprobe in der Wirkungsrichtung keinen eindeutigen Einfluß auf die Schwerbehindertenquote.

Abschließend soll noch auf ein methodisches Problem im Zusammenhang mit der Verwendung der Schwerbehindertenquote als abhängiger Variable eingegangen werden. Eliminiert man durch Einfügen einer Dummyvariablen "Schwerbehinderte beschäftigt" (0 keine Schwerbehinderten beschäftigt, 1 Schwerbehinderte beschäftigt) anstelle des Residuums aus der logistischen Regression den Einfluß von Betrieben ohne beschäftigte Schwerbehinderte<sup>158</sup>, dann hat die Betriebsgröße einen signifikant negativen Einfluß auf die Schwerbehindertenquote (vgl. Tabelle 36 und Tabelle 13 im Anhang). Durch das Herausfiltern von Betrieben ohne beschäftigte Schwerbehinderte fallen viele Kleinbetriebe mit einer Quote von Null aus der Analyse heraus, während viele kleine Betriebe mit einer hohen Quote in der Regressionsschätzung berücksichtigt werden. Kleine Betriebe mit beschäftigten Schwerbehinderten besitzen im Vergleich zu großen Betrieben eine höhere Istquote. So führt die Beschäftigung eines Schwerbehinderten in einem Unternehmen mit 20 Beschäftigten bereits zu einer Quote von 5%, d.h. alle Betriebe mit 20 Beschäftigten weisen bei Nichtberücksichtigung von Unternehmen ohne beschäftigten Schwerbehinderten eine Mindestquote von 5% auf. Die Quote reduziert sich mit einem Anstieg der Beschäftigtenzahl und einem beschäftigten Schwerbehinderten nur langsam.

Die Funktion des Schwerbehindertenanteils in Abhängigkeit von Beschäftigtenzahl und beschäftigten Schwerbehinderten ist für kleine Betriebe nicht stetig, d.h. erhöht sich bei gegebener Betriebsgröße die Zahl der beschäftigten Schwerbehinderten um eins, dann führt dies zu einem erheblichen Anstieg der Schwerbehindertenquote. Mit zunehmender Betriebsgröße wird die Steigerung der Schwerbehindertenquote geringer.

Die empirischen Analysen in den Abschnitten IV.4.4.2.2 und IV.4.4.3 verfolgen zwei Ziele: Zum einen sollte anhand der Schätzung der Arbeitsnachfragefunktion und der Determinanten der Lohnhöhe gezeigt werden, daß die Ergebnisse auf Basis der synthetischen Stichprobe einerseits aus neoklassischen Modellen abgeleiteten Implikationen nicht widersprechen und andererseits mit empirischen Befunden auf Basis realer Daten nicht im Widerspruch stehen. Zum anderen wurde durch die Schätzung einer Produktionsfunktion versucht, den aus der Theorie der

---

<sup>158</sup> Wenn die Regressionsschätzung nur für Betriebe mit beschäftigten Schwerbehinderten durchgeführt wird, resultiert das gleiche Ergebnis.

Internen Arbeitsmärkte bzw. aus moralökonomischen Ansätzen abgeleiteten Zusammenhang zwischen Organisationskapital und Bruttowertschöpfung empirisch zu testen sowie die Ergebnisse der Bestimmungsgründe der betrieblichen Schwerbehindertenbeschäftigung auf der Basis von Unternehmensdaten zu überprüfen.

Tabelle 35:Nachfrage nach schwerbehinderten Arbeitnehmern

Variable	B	SE B	Beta	T
Konzentrationsgrad	540,8116	50,4008	0,1830	10,730**
Residuum Schwerb.	6,0953	0,1226	0,5882	49,722**
Frauenanteil	-0,0053	0,0022	-0,0312	-2,369*
Beschäftigte	0,5012	0,0564	0,1206	8,880**
Facharbeiteranteil	0,0026	0,0027	0,0120	0,984
Ausländeranteil	-0,0087	0,0079	-0,0136	-1,089
Auszubildendenanteil	0,0132	0,0106	0,0168	1,243
Bruttowertschöpfung	-0,5697	0,2772	-0,0381	-2,055*
Lohn	0,7709	0,3060	0,0373	2,519*
Anlagevermögen	0,2062	0,2463	0,0178	0,837
Konstante	-2,9559	3,8704		-0,764
R <sup>2</sup> *100	39,972			
F-Wert	285,603**			

\*\*: $\alpha < 0,01$ ; \*: $\alpha < 0,05$

Die Resultate bezüglich der Arbeitsnachfrage sowie der Determinanten der Lohnhöhe stehen weitgehend in Einklang mit den theoretischen Konzepten und empirischen Analysen auf Basis echter Unternehmensdaten. Zu nicht eindeutigen Ergebnissen führt hingegen die Analyse der Determinanten der betrieblichen Schwerbehindertenbeschäftigung, auch wenn die Ergebnisse von Frick teilweise bestätigt werden konnten. Ursächlich für die im Gegensatz zu den Ergebnissen von Frick teilweise insignifikanten Schätzkoeffizienten können einerseits die oben skizzierten methodischen Probleme bezüglich der Schwerbehindertenquote und andererseits die nicht nur aufgrund der Berücksichtigung von "Unternehmen ohne beschäftigte Schwerbehinderte" wesentlich größere Varianz der Schwerbehindertenquote sein.

Ein positiver Einfluß aufeinander abgestimmter Rekrutierungs-, Qualifizierungs- und Entlohnungspraktiken interner Arbeitsmärkte, die zu einer rücksichtsvollen und langfristigen Nutzung der Arbeitskraft (Organisationskapital) führen und als deren Kuppelprodukt die (Weiter-)Beschäftigung von Schwerbehinderten angesehen werden kann, auf die Bruttowertschöpfung, konnte jedoch nicht direkt nachgewiesen werden, obwohl in dem Regressionsansatz die für die Bruttowert-

schöpfung wesentliche Größe, das Bruttoanlagevermögen, berücksichtigt wurde. Die im Modell berücksichtigten Variablen erklären nahezu 60% der Varianz der Bruttowertschöpfung pro Beschäftigtem. Auch ein Vergleich der Segregationsindizes zeigt, konträr zu den Ergebnissen von Frick, daß Schwerbehinderte bezüglich ihrer Beschäftigungschancen ähnlich diskriminiert sind wie Frauen und Ausländer.

Tabelle 36: Nachfrage nach schwerbehinderten Arbeitnehmern für Unternehmen mit beschäftigten Schwerbehinderten

Variable	B	SE B	Beta	T
Konzentrationsgrad	396,9500	50,9121	0,1343	7,797**
Sb. im Betrieb (Dummy)	5,9070	0,1216	0,6131	48,577**
Frauenanteil	-0,0059	0,0022	-0,0353	-2,659**
Beschäftigte	-0,2471	0,0590	-0,0590	-4,190**
Facharbeiteranteil	0,0020	0,0027	0,0091	0,741
Ausländeranteil	-0,0051	0,0080	-0,0080	-0,638
Auszubildendenanteil	0,0289	0,0107	0,0367	2,697**
Bruttowertschöpfung	-0,3568	0,2800	-0,0238	-1,276
Lohn	0,1783	0,3088	0,0086	0,577
Anlagevermögen	0,2600	0,2484	0,0225	1,047
Konstante	-0,9038	3,9033		-0,232
R <sup>2</sup> *100	38,956			
F-Wert	273,702**			

\*\*: $\alpha < 0,01$

\*: $\alpha < 0,05$

## **V. Zusammenfassung und Ausblick**

### **1. Eignung des Simulated Annealing zur Gewinnung synthetischer Unternehmensdaten**

Ausgehend von der Feststellung eines erheblichen Datendefizits auf Ebene der Unternehmen (Betriebe, Arbeitsstätten) von Seiten verschiedener Autoren und den Ergebnissen von Stafford (1986) und Hamermesh (1988), die die großen Fortschritte im Bereich der empirischen Forschung zum Arbeitsangebot auf ein Zusammenspiel von Theoriebildung, Methodenentwicklung und - in ganz entscheidendem Maße - von verfügbaren Daten zurückführen, erfolgte in Kapitel II eine exemplarische Bestandsaufnahme bisheriger Ansätze zur Beseitigung des Datenproblems auf Mikroebene. Fritsch (1990a, S. 40ff) sieht die wesentlichen Vorteile eines Ansatzes auf Mikroebene insbesondere darin, daß vorhandene Heterogenität entdeckt und untersucht werden kann, in der möglichen Überprüfung der Aussagefähigkeit von Aggregationsmerkmalen sowie in der Möglichkeit, Wirkungszusammenhänge besser aufdecken zu können. Die Analyse bisheriger Ansätze zur Lösung des Datenproblems auf Mikroebene zeigt, daß es keinen "Königsweg" zur Lösung des Datenproblems gibt. Mit der Bereitstellung eines Unternehmenspanels wäre jedoch ein wichtiger, aber auch sehr kostenintensiver Schritt zur Beseitigung des Datenproblems getan, sofern das Panel so konzipiert wäre, daß - ähnlich dem Sozio-ökonomischen Panel - eine Vielzahl von Forschungsfragen mit dem Panel beantwortet werden könnte und eine unmittelbare Datenweitergabe an alle Wissenschaftler sichergestellt ist.

Mit dem vom IAB derzeit durchgeführten Betriebspanel lassen sich sicherlich zukünftig eine ganze Reihe bisher ungeklärter Problemfelder bearbeiten. Die Beschränkung auf arbeitsmarktpolitische Fragen einerseits und fehlende "harte Daten" wie Umsatz, Wertschöpfung oder Anlagevermögen andererseits machen auch zukünftig die Erschließung oder Schaffung neuer Datenquellen notwendig, insbesondere dann, wenn der Zusammenhang zwischen interner Organisationsstruktur und Unternehmenserfolg auch nur annähernd erklärt werden soll. Betrachtet man nur die betriebliche Nachfrage und Nutzung von Arbeitskraft, dann kann die Effizienz verschiedener betrieblicher Strategien nur dann beantwortet werden, wenn Bedingungen herausgearbeitet werden, unter denen sich eine Strategie als vorteilhaft gegenüber einer anderen erweist. Die Effizienz eines Unternehmens kann sich langfristig nur an dessen ökonomischem Erfolg zeigen, der nicht zuletzt die Existenz des Unternehmens sicherstellt. Vorrangiges Ziel zukünftiger empirischer Analysen auf dem Forschungsfeld der Arbeitsnachfrage sollte daher die Analyse des Einflusses der betrieblichen Arbeitskräftenachfrage und der Nutzung von Arbeitskraft auf die Wertschöpfung oder den Gewinn des Unternehmens sein.

Auch wenn synthetische Unternehmensdaten die von Hamermesh (1988) idealtypisch formulierten Anforderungen an Daten nicht in vollem Umfang erfüllen, zeigt die Analyse bisheriger Ansätze zur Beseitigung des Datenproblems auf Unternehmensebene, daß in der Gewinnung synthetischer Daten ein vielversprechender und vor allem kostengünstiger Ansatz zur Beseitigung des Datenproblems liegt, wenn es gelingt, ein rationales Konzept für die Schätzung der unbekanntem Verteilung der Grundgesamtheit zu finden. Mit dem Konzept der "maximalen Entropie" steht ein Konstruktionsprinzip zur Verfügung, das einerseits gewünschte theoretische Eigenschaften besitzt und sich rational begründen läßt, andererseits basieren viele theoretische Verteilungen auf diesem Prinzip. Sind von einer Verteilung nur Erwartungswert und Varianz bekannt, dann resultiert als entropiemaximale Verteilung die Normalverteilung. Die Definition der Entropie stellt sicher, daß die Funktion ihren Maximalwert gerade dann erreicht, wenn die Verteilung mit der größten Unbestimmtheit, die Gleichverteilung, zugrunde liegt. In der Literatur wird dieses Prinzip als die "Leitvorstellung" zur Konstruktion synthetischer Daten bezeichnet.

Da bei der zu lösenden Aufgabe, der Konstruktion einer synthetischen Unternehmensstichprobe, eine große Zahl von Zellen an verschiedene Randbedingungen anzupassen ist, handelt es sich um ein Problem der kombinatorischen Optimierung, für dessen Lösung sich exakte Optimierungsverfahren nicht eignen. Mit dem Simulated Annealing steht ein leistungsfähiger Algorithmus zur Lösung derartiger Probleme zur Verfügung. Die theoretische Analyse des Algorithmus beweist dessen Konvergenz zum Minimum der Kostenfunktion, die die Abweichung der Randverteilungen auf Basis der synthetischen Stichprobe zu den vorgegebenen Randbedingungen mißt. Zudem konnte gezeigt werden, daß das Simulated Annealing unter gewissen Bedingungen entropiemaximale Lösungen erzeugt.

In umfangreichen Simulationsexperimenten konnte zum einen gezeigt werden, daß die programmierte Version die theoretisch postulierten Eigenschaften besitzt, und zum anderen konnten Parameterkonstellationen herausgearbeitet werden, mit denen die Wahrscheinlichkeit, die entropiemaximale Verteilung zu erreichen, zunimmt. Schließlich legt die Rekonstruktion einer realen Stichprobe die Vermutung nahe, daß vom Simulated Annealing erzeugte Stichproben sich nicht beliebig von realen Stichproben unterscheiden.

Im Mittelpunkt der vorliegenden Arbeit steht die generelle Prüfung, ob durch das Simulated Annealing erzeugte Unternehmensdaten sich an die Randbedingungen anpassen und mit diesen verträglich sind, d.h. die generelle Prüfung der Eignung des Verfahrens zur Konstruktion synthetischer Unternehmensdaten. In einem zweiten Schritt sollte untersucht werden, wie durch den Einsatz synthetischer Unternehmensdaten empirische Analysen im Zusammenhang mit einer substanzwissenschaftlichen Fragestellung durchgeführt werden können. Wie oben

bereits thematisiert, ist die Gegenüberstellung organisationsökonomischer Perspektiven und Implikationen mit dem Unternehmenserfolg dabei von besonderem Interesse. Aus diesem Grunde wird daher für die Konstruktion der synthetischen Stichprobe auf drei Datenquellen zurückgegriffen, deren Erfassungsgenauigkeit als sehr gut angesehen werden kann und die eine Verknüpfung von organisationsökonomisch relevanten Variablen mit Meßgrößen des Unternehmenserfolgs ermöglichen: die Arbeitsstättenzählung von 1987, die Zeitreihen des Bruttoanlagevermögens und der Bruttowertschöpfung des Deutschen Instituts für Wirtschaftsforschung sowie die Auswertungen der Arbeitgeberanzeigen gemäß § 13 Abs. 2 SchwbG durch die Bundesanstalt für Arbeit. Durch Anwendung des Simulated Annealing lassen sich diese unterschiedlichen Aggregatdaten über eine oder mehrere Variablen, die in allen Datensätzen vorhanden sind, miteinander verknüpfen.

Die Qualität der synthetischen Stichprobe läßt sich nicht durch einen Vergleich mit einer realen Stichprobe beurteilen. Als Qualitätsindikatoren können jedoch die Anpassung an die Randverteilungen sowie der Vergleich von Ergebnissen auf Basis von synthetischen Stichprobe mit den Ergebnissen auf Basis echter Daten herangezogen werden. Die Anpassung an die Randverteilungen gelingt in allen mit unterschiedlichen Startwerten des Zufallszahlengenerators erzeugten Stichproben sehr gut. Der Vergleich mit Ergebnissen auf Basis realer Stichproben konnte zeigen, daß das Betriebsgrößenlohndifferential auch bei Kontrolle einer Vielzahl anderer Einflußfaktoren nicht verschwindet. Zudem stehen die Ergebnisse für die Einflußfaktoren der betrieblichen Arbeitsnachfrage bzw. der Lohnhöhe nicht im Widerspruch zu theoretisch abgeleiteten Implikationen sowie bisherigen empirischen Ergebnissen.

Sowohl die theoretische Analyse des Simulated Annealing und die Simulationsexperimente als auch die empirischen Ergebnisse zu den Determinanten der Lohnhöhe zeigen die Eignung des Simulated Annealing zur Konstruktion synthetischer Unternehmensdaten. Eines der überzeugendsten Ergebnisse, die das Simulated Annealing für die Konstruktion synthetischer Unternehmensdaten als geeignet erscheinen lassen und das möglicherweise sogar für die Realitätsnähe so erzeugter Daten spricht, ist die nahezu exakte Abbildung von Betrieben ohne beschäftigte Schwerbehinderte in allen sechs Stichproben, obwohl der Anteil dieser Betriebe nicht durch eine entsprechende Randbedingung vorgegeben wurde.

## **2. Wertschöpfung und Schwerbehindertenbeschäftigung: Zur empirischen Evidenz organisationsökonomischer Perspektiven auf Basis synthetischer Unternehmensdaten**

Unter Rückgriff auf die Theorie interner Arbeitsmärkte kommt Frick (1992, S. 243ff) zu dem Ergebnis, daß stabile Beschäftigungsverhältnisse, innerbetriebliche

Mobilitätsketten und an der Betriebszugehörigkeitsdauer orientierte Entlohnungssysteme die Entstehung langfristig "effizienter" Kooperations- und Vertrauensbeziehungen mit daraus resultierenden Kooperationsrenten erleichtern, da die Minimierung der Arbeitskosten nicht nur eine Minimierung der Rekrutierungs-, Qualifizierungs- und Lohn- bzw. Gehaltskosten, sondern auch eine Minimierung der Motivations- und Kontrollkosten bedeutet. In der (Weiter)Beschäftigung von Schwerbehinderten sieht der Autor ein Kuppelprodukt der Arbeitskräftestrategie der Internalisierung. In seiner empirischen Analyse auf Basis von Branchendaten kommt Frick zu dem Ergebnis, daß eine Internalisierungsstrategie (gemessen durch die Stabilitätsrate und Fluktuationsrate) tatsächlich zu einem größeren Anteil beschäftigter Schwerbehinderter führt.

Läßt man die Schwerbehindertenquote folglich als Indikator für eine Internalisierungsstrategie gelten (Bestimmtheitsmaß der Regressionsfunktion 76,9%, erklärende Variable Fluktuationsrate und Stabilitätsrate), dann müßte mit zunehmendem Schwerbehindertenanteil eine größere Bruttowertschöpfung zu beobachten sein. In dem vorgestellten Regressionsmodell zur Erklärung der Bruttowertschöpfung kann für das Verarbeitende Gewerbe, selbst bei Kontrolle des Anlagevermögens und der Berücksichtigung einiger von Frick zur Erklärung der Fluktuationsrate und Stabilitätsrate herangezogener Variablen, kein positiver Einfluß der Schwerbehindertenquote auf die Bruttowertschöpfung pro Arbeitnehmer nachgewiesen werden. Vielmehr ist der Einfluß der Schwerbehindertenquote auf die Bruttowertschöpfung in allen sechs Stichproben signifikant negativ.

Wenn auch die Ergebnisse auf Basis einer synthetischen Stichprobe geschätzt werden, deuten die Ergebnisse dennoch darauf hin, daß der gewählte theoretische Ansatz möglicherweise zu kurz greift. Aus der Signalfunktion der (Weiter)Beschäftigung von Schwerbehinderten auf eine höhere Leistungsbereitschaft und Loyalität der Beschäftigten zu schließen, impliziert, daß nichtbehinderte, bzw. gesundheitlich beeinträchtigte Arbeitnehmer das individuelle Risiko der Schwerbehinderung auch für sich selbst erkennen. Nicht zuletzt die Einführung der Pflegeversicherung (und anderer Sozialversicherungen) wird damit begründet, daß gesundheitliche Risiken systematisch unterschätzt werden. Es bedarf daher einer weitergehenden Begründung, warum gerade das Risiko der Schwerbehinderung von nichtbehinderten Arbeitnehmern für sich selbst antizipiert werden sollte, denn nur dann besäße die (Weiter)Beschäftigung von Schwerbehinderten die postulierte Signalfunktion.

Wenn Frick (1992, S. 245) auch darauf abstellt, daß eine gezielte Ausgliederung oder Diskriminierung Schwerbehinderter für den Fall wenig wahrscheinlich ist, daß eine (Weiter-)Beschäftigung den betrieblichen Interessen an einer Minimierung der Arbeitskosten als auch den Reziprozitätserwartungen der Belegschaft entspricht, kann jedoch aufgrund der Ergebnisse auf Basis der synthetischen Stichprobe bezweifelt werden, ob die (Weiter-)Beschäftigung von Schwerbehin-

derten in jedem Falle ökonomisch effizient ist. Wenn die Nichtbeschäftigung von Schwerbehinderten jedoch ökonomische Vorteile zur Folge hat und die Beschäftigung von Schwerbehinderten erklärtes sozialpolitisches Ziel ist, dann besteht die Aufgabe der Sozialpolitik darin, die Lasten zur Erreichung dieses sozialpolitischen Zieles gerechter zu verteilen, d.h. die Nichterfüllung der Pflichtquote wäre mit höheren Sanktionen zu belegen.

Auf erhebliche methodische Probleme im Zusammenhang mit der Schwerbehindertenquote als abhängiger Variable in Regressions-schätzungen kann aus der Schätzung des Schwerbehindertenanteils bei Berücksichtigung der Dummyvariable (Schwerbehinderte beschäftigt? ja/nein, vgl. Tabelle 36) geschlossen werden. Beschäftigen kleine und mittlere Unternehmen überhaupt einen Schwerbehinderten, dann weisen sie gleichzeitig einen großen Schwerbehindertenanteil auf<sup>159</sup>. Neben der wesentlich größeren Streuung des Schwerbehindertenanteils (als abhängiger Variable) und der unabhängigen Variablen auf Unternehmensebene kann die Unstetigkeit der Schwerbehindertenquote für kleine und mittlere Unternehmen Ursache für die im Vergleich zu Frick doch sehr uneinheitlichen Ergebnissen bei den Bestimmungsgründen der betrieblichen Schwerbehindertenbeschäftigung sein.

Angesichts der methodischen Probleme (Unstetigkeit der Schwerbehindertenquote bei kleineren und mittleren Betrieben) kann bezweifelt werden, ob sich die Schwerbehindertenquote in ökonometrischen Schätzmodellen auf Basis von Unternehmensdaten als abhängige Variable eignet. Schätzungen auf Basis von Branchendaten führen nicht zu derartigen methodischen Problemen, da die Schwerbehindertenquote in einer Branche durch Division der beschäftigten Schwerbehinderten mit den Beschäftigten insgesamt berechnet wird, so daß die Unstetigkeit zwar beseitigt wird, aber auch eine erhebliche Varianzreduktion die Folge ist. Hält man an der Schwerbehindertenquote pro Unternehmen, trotz aller methodischen Probleme, als sozialpolitischem Indikator zur Steuerung der Integration Schwerbehinderter in das Erwerbsleben fest, dann ist jedoch nicht die Schwerbehindertenquote auf Branchenebene von Interesse, sondern potentielle Beschäftigungsreserven sind bei Nichterfüllern der gesetzlichen Pflichtquote auf Unternehmensebene vorhanden. Dies zeigt sich vor allem darin, daß die Istquote auf Unternehmensebene, gegeben eine Behindertenquote auf Branchenebene von ca. 5,6%, auf Unternehmensebene nur ca. 4,4% beträgt, d.h. die Istquote auf Unternehmensebene zeigt wesentlich größere Beschäftigungsreserven. Zukünftigen Analysen zu den Bestimmungsgründen der betrieblichen Schwerbehindertenbeschäftigung, auf deren Basis sozialpolitische Entscheidungen getroffen werden, sollten daher Unternehmensdaten zugrundeliegen, da nur dann Aussagen über die

---

<sup>159</sup> Vgl. zu dieser Problematik auch Sadowski et. al. 1992, S. 185f.

Ursachen der Nichterfüllung oder Übererfüllung der Pflichtquote pro Unternehmen möglich sind.

### **3. Weitere empirische Implikationen auf Basis synthetischer Unternehmensdaten**

In originär erhobenen Stichproben gelingt es selten, harte Variable wie Umsatz, Anlagevermögen, Investitionen oder sogar den Gewinn eines Unternehmens zu erheben. Die Verknüpfung von Variablen, die eher die interne Organisationsstruktur eines Unternehmens kennzeichnen und damit möglicherweise auf das Vorhandensein einer Arbeitskräftestrategie der Internalisierung hindeuten sowie Meßgrößen des Unternehmenserfolgs und der Kapitalstruktur wird durch die Anwendung des Simulated Annealing möglich. Die synthetische Stichprobe kann an eine Vielzahl von Randverteilungen, aber auch andere Größen (z.B. Variablenbeziehungen) angepaßt werden. In der hier durchgeführten Anwendung wurden nur Randverteilungen aus drei verschiedenen Erhebungen miteinander verknüpft, was dennoch zu einem Datensatz führte, der die Größen "durchschnittliches Anlagevermögen" und "Bruttowertschöpfung pro Beschäftigtem" neben einigen anderen Variablen enthält.

Die geschätzte Produktionsfunktion mit der durchschnittlichen Bruttowertschöpfung als abhängiger Variable (vgl. Tabelle 31 und Tabelle 8 im Anhang) zeigt dann auch zum Teil sehr kontraintuitive Ergebnisse. Der signifikant negative Einfluß des Facharbeiteranteils, auch bei Berücksichtigung des Bruttoanlagevermögens, in allen sechs Stichproben ist sicherlich sehr überraschend und steht zu den meisten theoretischen Modellen und empirischen Befunden im Widerspruch. Dieser Zusammenhang kann möglicherweise dadurch erklärt werden, daß Bruttoanlagevermögen und Facharbeiteranteil von einander abhängen. Je größer das Bruttoanlagevermögen im Unternehmen, desto mehr qualifizierte Arbeitskräfte sind zu dessen Bedienung erforderlich. Die Schätzung einer Faktornachfragefunktion mit dem Facharbeiteranteil als abhängiger Variable unterstützt diese These, da der Facharbeiteranteil signifikant positiv vom Bruttoanlagevermögen pro Beschäftigten abhängt. Aufgrund der geringen Varianzaufklärung (zwischen 4% u. 12%) in den einzelnen Stichproben ist nur eine sehr vorsichtige Interpretation dieses Zusammenhangs möglich.

Die geschätzte Produktionsfunktion, in der auch der Frauenanteil als unabhängige Variable berücksichtigt wurde, läßt auf einen leicht negativen Einfluß dieser Variable auf die Bruttowertschöpfung mit einem geringen relativen Erklärungsbeitrag schließen. Es kann unterstellt werden, daß mit dem Frauenanteil der Anteil der Teilzeitbeschäftigten im Unternehmen steigt, so daß der negative Einfluß möglicherweise auf einen rein rechnerischen Effekt zurückzuführen ist. Teilzeitbeschäftigte gehen in die Berechnung des Bruttowertschöpfung pro Beschäftig-

tem mit gleicher Gewichtung ein wie Vollzeitbeschäftigte, so daß eine niedrigere durchschnittliche Wertschöpfung bei höherem Anteil an Teilzeitbeschäftigten resultiert. Dieser rein rechnerische Effekt wirkt auf ähnliche Weise auf die Höhe des Durchschnittslohns pro Beschäftigten. Betrachtet man die Regressionsergebnisse zu den Bestimmungsgründen des Durchschnittslohn pro Beschäftigten, zeigt sich der aufgrund des rechnerischen Effektes erwartete negative Zusammenhang. Dabei ist jedoch der relative hohe Erklärungsbeitrag des Frauenanteils sowie der deutlich niedrigere Wert des Regressionskoeffizienten (0,0006 auf die Bruttowertschöpfung und -0,0027 auf den Durchschnittslohn) überraschend. Bei recht hohen Varianzaufklärungen beider Schätzfunktionen deutet dieser Befund auf eine ausgeprägte Einkommensdiskriminierung von Frauen hin, die nicht ausschließlich durch Produktivitätsunterschiede erklärt werden kann.

#### **4. Ausblick: Zukünftige Forschung auf Basis synthetischer Unternehmensdaten**

Es kann sicherlich davon ausgegangen werden, daß als Kritik an der vorliegenden Arbeit der Einwand erhoben wird, die konstruierte Stichprobe stelle zwar eine "nette Spielerei" dar, eigne sich aber für die Produktion echten empirischen Wissens kaum, da die Variablenbeziehungen bis auf die Restriktionen durch die vorgegebenen Randbedingungen beliebig seien. Die theoretische Analyse, das der Konstruktion zugrundeliegenden "Prinzips der maximalen Entropie" sowie die empirischen Ergebnisse auf Basis der synthetischen Stichprobe verdeutlichen jedoch zum einen die Rationalität des Konstruktionsprinzips und zum anderen die "Realitätsnähe" der konstruierten Daten. Die Ausführungen in Abschnitt III.7 zeigen, daß auch bei der Gewichtung realer Stichproben, wenn überhaupt auf eine Erhaltung der Strukturinformationen in der erhobenen Stichprobe Wert gelegt wird, das Prinzip des minimalen Informationsverlustes, welches eng mit dem Entropieprinzip verbunden ist, zur Anwendung kommt. Rothe (1990) zeigt am Beispiel des ALLBUS, wie durch Gewichtungsfaktoren, die nur auf einem Teil der erfaßten Variablen berechnet werden, erhebliche Verzerrungen bei den nicht in die Gewichtungsfaktoren einbezogenen Variablen resultieren. Die Gewichtung von Stichproben entfernt daher möglicherweise die erhobene Stichprobe von den Strukturen der Grundgesamtheit, während mit der Konstruktion der synthetischen Stichprobe gerade versucht wird, sich diesen anzunähern.

Wenn die angeführten Argumente als noch nicht ausreichend für die Eignung synthetischer Stichproben in der empirischen Forschung angesehen werden, dann sollten auch die größten Skeptiker anerkennen, daß sich auf Basis synthetischer Daten zumindest Erkenntnisse darüber gewinnen lassen, welche möglichen Strukturen sich hinter Aggregatdaten verbergen können. Die Interpretation von Branchendaten als repräsentatives Unternehmen ist daher ebenso begründungsbedürftig wie die empirische Arbeit mit synthetischen Unternehmensdaten.

Die Annäherung an reale Strukturen ist von der Qualität und Quantität der vorhandenen Aggregatinformationen abhängig, d.h. je mehr Variablen insgesamt bzw. gemeinsam in verschiedenen Aggregatdatensätzen vorhanden sind. Wenn eine Variable in mehreren Aggregatdatensätzen vorhanden ist, dann ist eine Verknüpfung verschiedener Datensätze über diese gemeinsame Variable möglich. Es setzt jedoch auch voraus, daß unterschiedliche Aggregatdatensätze auch miteinander verknüpft werden können. Bereits bei der Berücksichtigung von nur drei Datenquellen aus der Bundesrepublik Deutschland sind umfangreiche Umgliederungen notwendig, um deren Vergleichbarkeit herstellen. Die Darstellung der von amtlichen Stellen erhobenen Daten und deren Veröffentlichungsform verdeutlicht, daß eine Verknüpfung aufgrund unterschiedlicher Gruppierungsmerkmale und Variablenabgrenzungen nur schwer oder überhaupt nicht möglich ist. Wenn auch den Informationsinteressen unterschiedlicher Adressatengruppen amtlicher Datenerhebungen Rechnung getragen werden muß, wäre mit einer Vereinheitlichung oder Vergleichbarkeit von verschiedenen Datenquellen ein großer Schritt für die Forschung und damit auch für die Politikberatung getan. Diese Daten stünden dann für die Disaggregation mittels verschiedener Optimierungsverfahren zur Verfügung, so daß auf Basis dieser Daten synthetische Stichproben erzeugt werden könnten, mit denen zumindest Aussagen über die sich hinter den Aggregaten verbergenden Zusammenhänge getroffen werden könnten.

Die hier vorgestellten Ergebnisse basieren auf einem synthetischen Querschnittsdatsatz. Viele theoretische Modelle der Arbeitsnachfrage sind jedoch als dynamische Modelle formuliert, so daß sich Querschnittsdaten zur empirischen Überprüfung von Implikationen aus diesen Modellen nur bedingt eignen. Mit dem vorgestellten Ansatz erscheint es jedoch auch möglich, Längsschnittsdatsätze zu erzeugen. Dies setzt jedoch voraus, daß auf Ebene der Aggregate (z.B. Branchen und/oder Betriebsgrößenklassen) Übergangsraten bekannt sind und veröffentlicht werden. Mittels des Simulated Annealing könnten diese aggregierten Übergangsraten disaggregiert und auf Unternehmensebene berechnet werden. Als Randbedingungen würden dann nicht Bestandsdaten zugrundegelegt, sondern die aggregierten Übergänge, beispielsweise von einer Betriebsgrößenklasse in eine andere aufgrund von Unternehmenswachstum oder -schrumpfung. Auf ähnliche Weise ließe sich die Veränderungsrate von Facharbeiteranteil, Frauenanteil, Schwerbehindertenanteil, Anlagevermögen, Wertschöpfung usw. in die synthetische Stichprobe hineinsimulieren. Ein derartiges komplexes Vorhaben ist jedoch nicht mehr in einem einzelnen Dissertationsprojekt zu realisieren, sondern würde vor allem im mathematischen, statistischen und programmiertechnischen Bereich größere Ressourcen erfordern als dem Autor dieser Arbeit zur Verfügung stehen.

Ist bereits die Verfügbarkeit von Unternehmensdaten auf nationaler Ebene als dürftig zu bezeichnen, zeigt sich für die international vergleichende empirische Forschung ein noch düsteres Bild. "Sind schon nationale Betriebsdaten selten, so sind international vergleichbare Daten eine wertvolle Rarität. Auf Basis des

vorliegenden Datenmaterials der amtlichen Statistik sowie den wenigen vorhandenen Betriebsdatensätzen kann personalpolitisches Verhalten auf (grenznahen) regionalen Arbeitsmärkten nicht explizit betrachtet werden" (Backes-Gellner/Böck/Wienecke 1994, S. 6).

Die oben bereits dargestellten Probleme auf nationaler Ebene hinsichtlich der unterschiedlichen Berichtssysteme (Periodizität, Erhebungsmerkmale, Erhebungsräume, sowie Erhebungszeitpunkte und -zeiträume) treten auf internationaler Ebene noch in stärkerem Maße auf. Von EUROSTAT wurde auf internationaler Ebene versucht, international vergleichbare Daten auf regionaler Ebene zusammenzustellen. Backes-Gellner et al. (1994) sehen jedoch in dem resultierenden Ergebnis oft nur eine Vielzahl von Umrechnungen, Aggregationen und Kompromissen, so daß bei der Interpretation dieser Daten erhebliche Vorsicht geboten bleibt. Die Autoren wählen aufgrund dieser Probleme den Weg einer eigenen Erhebung, die jedoch aufgrund begrenzter finanzieller und personeller Kapazitäten weder als Panelstudie konzipiert ist noch große Fallzahlen anstrebt, sondern die eher als ein Datensatz aus detaillierten Fallstudien anzusehen ist. Sie halten detaillierte Fallstudien für geeignet, um erste Erfahrungen mit international vergleichenden Betriebsbefragungen zu sammeln.

Auch wenn die international zur Verfügung stehenden Aggregatdatensätze mit Vorsicht zu interpretieren und zu beurteilen sind, könnten sie zur Konstruktion einer synthetischen Stichprobe genutzt werden. Wegen der im Vergleich zu nationalen Aggregatdatensätzen geringeren Vergleichbarkeit ist es sicherlich unangebracht, diese synthetische Unternehmensstichprobe als Abbildung der internationalen Unternehmensgrundgesamtheit zu begreifen. Es wären aber auf Basis von Unternehmensdaten empirische Analysen möglich, die erste Einblicke in Wirkungszusammenhänge im internationalen Vergleich erlauben. Derartige Ergebnisse ließen sich vielleicht zur Konzeption einer internationalen Studie nutzen, da auf Basis einer relativ großen Stichprobe tatsächlich vorhandene Zusammenhänge besser identifiziert werden können. Die Erkenntnisse auf Basis der synthetischen Stichprobe für die Konzeption international vergleichender Studien sind möglicherweise als komplementär zu den Erkenntnissen aus den detaillierten Fallstudien im Rahmen von QUIPPE (Quinter Studie zur Praxis der Personalpolitik in Europa) anzusehen.



**ANHANG**

Tabelle 1: Summe Beschäftigte in den Größenklassen	192
Tabelle 2: Summe Umsatz in den Größenklassen	193
Tabelle 3: Summe Lohn in den Größenklassen	193
Tabelle 4: Anzahl der Betriebe nach Lohn und Beschäftigten	194
Tabelle 5: Anzahl der Betriebe nach Umsatz und Beschäftigten	195
Tabelle 6: Mittelwerte und Standardabweichungen der Variablen „Bruttowertschöpfung, Bruttoanlagevermögen und Beschäftigte“ sowie der logarithmierten Werte	196
Tabelle 7: Cobb-Douglas-Produktionsfunktion für zwei Inputfaktoren	197
Tabelle 8: Produktionsfunktion mit differenziertem Arbeitsinput	198
Tabelle 9: Arbeitsnachfragefunktion für das Zwei-Faktoren-Modell	200
Tabelle 10: Determinanten des Durchschnittslohns pro Beschäftigten	201
Tabelle 11: Arbeitsnachfragefunktion mit differenzierten erklärenden Variablen	203
Tabelle 12: Nachfrage nach schwerbehinderten Arbeitnehmern	205
Tabelle 13: Nachfrage nach schwerbehinderten Arbeitnehmern für Unternehmen mit beschäftigten Schwerbehinderten	207

Tabelle 1: Summe Beschäftigte in den Größenklassen

Größenklasse	Beschäftigte
unter 100	906
100 b. u.500	26.761
500 b. u. 1.000	38.934
1.000 b. u. 2.000	79.505
2.000 b. u. 5.000	157.185
5.000 b. u. 10.000	249.682
10.000 b. u. 25.000	459.161
25.000 u. mehr	3.182.338

Tabelle 2: Summe Umsatz in den Größenklassen

Größenklasse	Umsatz
u. 10 Mio.	69,25
10 b. u. 50 Mio.	1.315,06
50 b. u. 100 Mio.	3.356,68
100 b. u. 200 Mio.	8.197,34
200 b. u. 500 Mio.	20.027,91
500 b. u. 1000 Mio.	30.819,97
1000 b. u. 5000 Mio.	142.909,67
ü. 5000 Mio.	769.901,72





Tabelle 6: Mittelwerte und Standardabweichungen der Variablen „Bruttowertschöpfung, Bruttoanlagevermögen und Beschäftigte“ sowie der logarithmierten Werte

Variable	Startwert					
	1		2		3	
	Mittelwert	Stdabw.	Mittelwert	Stdabw.	Mittelwert	Stdabw.
Bruttowertschöpfung	51.323,94	15.243,54	51.308,04	15.194,78	51.300,45	15.224,48
Bruttowertschöpfung (logarithmiert)	10,81	0,28	10,81	0,28	10,81	0,28
Anlagevermögen	103.653,73	46.528,71	103.627,42	46.577,62	103.677,86	46.649,03
Anlagevermögen (logarithmiert)	11,47	0,36	11,47	0,36	11,47	0,36
Beschäftigte	170,78	1.316,04	170,78	1.316,04	170,78	1.316,04
Beschäftigte (logarithmiert)	4,02	1,01	4,02	1,01	4,02	1,01

Variable	Startwert					
	4		5			
	Mittelwert	Stdabw.	Mittelwert	Stdabw.	Mittelwert	Stdabw.
Bruttowertschöpfung	51.328,73	15.224,70	51.337,25	15.245,55		
Bruttowertschöpfung (logarithmiert)	10,81	0,28	10,81	0,28		
Anlagevermögen	103.644,48	46.546,93	103.641,78	46.571,70		
Anlagevermögen (logarithmiert)	11,47	0,36	11,47	0,36		
Beschäftigte	170,78	1.316,04	170,78	1.316,04		
Beschäftigte (logarithmiert)	4,02	1,01	4,02	1,01		

Tabelle 7: Cobb-Douglas-Produktionsfunktion für zwei Inputfaktoren

Variable	Startwert											
	1				2				3			
	B	SE B	Beta	T	B	SE B	Beta	T	B	SE B	Beta	T
Beschäftigte	0,0560	0,0029	0,2012	19,540**	0,0552	0,0029	0,1990	19,322**	0,0564	0,0029	0,2027	19,668**
Anlagevermögen	0,5331	0,0080	0,6897	66,971**	0,5315	0,0079	0,6904	67,033**	0,5312	0,0080	0,6890	66,810**
Konstante	4,4639	0,0906		49,253**	4,4856	0,0903		49,695**	4,4838	0,0905		49,532**
R <sup>2</sup> *100	55,175				55,171				55,107			
F-Wert	2.644,597**				2.644,128**				2.637,353**			

Variable	Startwert											
	4				5							
	B	SE B	Beta	T	B	SE B	Beta	T				
Beschäftigte	0,0562	0,0029	0,2024	19,736**	0,0570	0,0029	0,2048	19,911**				
Anlagevermögen	0,5341	0,0079	0,6919	67,463**	0,5324	0,0079	0,6890	66,987**				
Konstante	4,4522	0,0901		49,397**	4,4684	0,0905		49,380**				
R <sup>2</sup> *100	55,549				55,280							
F-Wert	2.684,948**				2.655,787**							

\*\*: $\alpha < 0,01$

\*: $\alpha < 0,05$

Tabelle 8: Produktionsfunktion mit differenziertem Arbeitsinput

Variable	Startwert											
	1				2				3			
	B	SE B	Beta	T	B	SE B	Beta	T	B	SE B	Beta	T
Lohn	0,2274	0,0161	0,1638	14,125**	0,2279	0,1652	0,1652	14,242**	0,1805	0,0168	0,1295	10,724**
Anlagevermögen	0,5059	0,0079	0,6544	63,665**	0,5130	0,6664	0,6664	65,762**	0,5033	0,0078	0,6524	64,256**
Facharbeiteranteil	-0,0014	0,0001	-0,0996	-10,081**	-0,0010	0,0001	-0,0720	-7,259**	-0,0009	0,0001	-0,0601	-6,139**
Schwerbehindertenanteil	-0,0024	0,0006	-0,0373	-3,815**	-0,0014	0,0006	-0,0208	-2,115*	-0,0014	0,0006	-0,0215	-2,188*
Ausländeranteil	-0,0021	0,0004	-0,0482	-4,822**	-0,0023	0,0004	-0,0568	-5,592**	-0,0018	0,0004	-0,0430	-4,302**
Auszubildendenanteil	0,0086	0,0005	0,1655	16,511**	0,0070	0,0005	0,1343	13,404**	0,0080	0,0005	0,1441	14,053**
Frauenanteil	-0,0003	0,0001	-0,0202	-1,916	0,0001	0,0001	0,0064	0,608	-0,0008	0,0001	-0,0677	-6,189**
Beschäftigte	0,0436	0,0030	0,1567	14,569**	0,0398	0,0030	0,1435	13,088**	0,0465	0,0031	0,1671	15,204**
Konstante	2,4739	0,1833		13,494**	2,3877	0,1859		12,843**	2,9804	0,1954		15,249**
R <sup>2</sup> *100	60,184				59,046				59,306			
F-Wert	810,748**				773,312**				781,692**			

Fortsetzung Tabelle 8

Variable	Startwert									
	4					5				
	B	SE B	Beta	T	B	SE B	Beta	T		
Lohn	0,1873	0,0160	0,1350	11,710**	0,2015	0,0162	0,1450	12,406**		
Anlagevermögen	0,5120	0,0079	0,6631	64,520**	0,5119	0,0078	0,6624	65,868**		
Facharbeiteranteil	-0,0010	0,0001	-0,0679	-6,997**	-0,0011	0,0001	-0,0762	-7,705**		
Schwerbehindertenanteil	-0,0031	0,0006	-0,0493	-5,010**	-0,0022	0,0006	-0,0342	-3,486**		
Ausländeranteil	-0,0020	0,0004	-0,0467	-4,701**	-0,0021	0,0004	-0,0479	-4,790**		
Auszubildendenanteil	0,0080	0,0005	0,1524	15,174**	0,0075	0,0005	0,1398	13,934**		
Frauenanteil	-0,0006	0,0001	-0,0478	-4,556**	-0,0003	0,0001	-0,0297	-2,814**		
Beschäftigte	0,0434	0,0030	0,1562	14,461**	0,0451	0,0031	0,1620	14,720**		
Konstante	2,8248	0,1828	0,1562	15,451**	2,6695	0,1882	0,1620	14,182**		
R <sup>2</sup> *100	60,085					59,370				
F-Wert	807,429**					783760**				

\*\*: $\alpha < 0,01$ \*: $\alpha < 0,05$

Tabelle 9: Arbeitsnachfragefunktion für das Zwei-Faktoren-Modell

Variable	Startwert											
	1				2				3			
	B	SE B	Beta	T	B	SE B	Beta	T	B	SE B	Beta	T
Wertschöpfung	0,7716	0,0498	0,2147	15,506**	0,7710	0,0500	0,2140	15,428**	0,7690	0,0498	0,2139	26,235**
Lohn	1,8195	0,0691	0,3648	26,350**	1,7960	0,0689	0,3614	26,057**	1,8230	0,0695	0,3638	15,430**
Konstante	-23,3239	0,8062		-28,931**	-23,0711	0,8067		-28,600**	-23,3306	0,8068		-28,918**
R <sup>2</sup> 100	21,132				20,806				21,111			
F-Wert	575,668**				564,454**				574,952**			

Variable	Startwert											
	4				5							
	B	SE B	Beta	T	B	SE B	Beta	T				
Wertschöpfung	0,7760	0,0498	0,2156	15,575**	0,7887	0,0496	0,2195	15,905**				
Lohn	1,8178	0,0691	0,3641	26,297**	1,8283	0,0690	0,3660	26,515**				
Konstante	-23,3527	0,8064		-28960**	-23,6003	0,8069		-29248**				
R <sup>2</sup> 100	21,157				21,434							
F-Wert	576,528**				586,152**							

\*\*: $\alpha < 0,01$

\*: $\alpha < 0,05$

Tabelle 10: Determinanten des Durchschnittslohns pro Beschäftigten

Variable	Startwert											
	1				2				3			
	B	SE B	Beta	T	B	SE B	Beta	T	B	SE B	Beta	T
Kapitalproduktivität	0,0017	0,0002	0,1016	7,888**	0,0023	0,0002	0,1381	10,820**	0,0018	0,0002	0,1120	9,021**
Auszubildendenanteil	-0,0069	0,0005	-0,1851	-14,300**	-0,0065	0,0005	-0,1710	-13,344**	-0,0055	0,0005	-0,1450	-11,275**
Schwerbehindertenanteil	0,0009	0,0005	0,0186	1,479	0,0006	0,0006	0,0131	1,040	0,0004	0,0005	0,0081	0,661
Ausländeranteil	0,0012	0,0004	0,0384	2,981**	0,0006	0,0003	0,0199	1,524	-0,0001	0,0004	-0,0029	-0,230
Facharbeiteranteil	0,0009	0,0001	0,0897	7,022**	0,0009	0,0001	0,0834	6,556**	0,0004	0,0001	0,0415	3,392**
Beschäftigte	0,0633	0,0027	0,3157	23,739**	0,0680	0,0027	0,3378	25,279**	0,0687	0,0026	0,3450	26,439**
Frauenanteil	-0,0027	0,0001	-0,3002	-23,534**	-0,0026	0,0001	-0,2955	-23,266**	-0,0030	0,0001	-0,3688	-29,596**
Bruttowertschöpfung	0,1234	0,0100	0,1713	12,383**	0,0982	0,0100	0,1354	9,911**	0,0657	0,0097	0,0916	6,784**
Konstante	8,8507	0,1055		83,916**	9,0751	0,1052		86,235**	9,4780	0,1028		92,236**
R <sup>2</sup> 100		34,324				32,679				36,934		
F-Wert		280,320**				260,365**				313,982**		

Fortsetzung Tabelle 10

Variable	4				Startwert				S			
	B	SE.B	Beta	T	B	SE.B	Beta	T	B	SE.B	Beta	T
Kapitalproduktivität	0,0014	0,0002	0,0846	6,431**	0,0021	0,0002	0,1256	9,886**				
Auszubildendenanteil	-0,0057	0,0005	-0,1499	-11,336**	-0,0061	0,0005	-0,1567	-12,156**				
Schwerbehindertenanteil	0,0010	0,0006	0,0225	1,750	1,3851	0,0005	0,0029	0,233				
Ausländeranteil	0,0007	0,0004	0,0222	1,710	0,0010	0,0004	0,0315	2,446*				
Facharbeiteranteil	0,0005	0,0001	0,0476	3,748**	0,0008	0,0001	0,0823	6,481**				
Beschäftigte	0,0701	0,0027	0,3499	26,169**	0,0712	0,0027	0,3557	26,582**				
Frauenanteil	-0,0027	0,0001	-0,3134	-24,379**	-0,0026	0,0001	-0,3065	-24,138**				
Bruttowertschöpfung	0,1035	0,0101	0,1437	10,249**	0,0820	0,0098	0,1140	8,329**				
Konstante	9,0682	0,1066		85,049	9,2486	0,1044		88,574**				
R <sup>2</sup> *100	32,298				33,319							
F-Wert	255,886**				268,012**							

\*\*: $\alpha < 0,01$ \*: $\alpha < 0,05$

Tabelle 11: Arbeitsnachfragefunktion mit differenzierten erklärenden Variablen

Variable	Startwert											
	1				2				3			
	B	SE.B	Beta	T	B	SE.B	Beta	T	B	SE.B	Beta'	T
Lohn	1,8852	0,0765	0,3779	24,629**	1,9417	0,0741	0,3907	26,203**	2,0785	0,0761	0,4148	27,296**
Ausländeranteil	0,0126	0,0022	0,0804	5,827**	0,0129	0,0020	0,0871	6,299**	0,0060	0,0020	0,0371	2,742**
Schwerbehindertenanteil	0,0142	0,0031	0,0610	4,539**	0,0199	0,0032	0,0837	6,253**	0,0208	0,0031	0,0877	6,621**
Facharbeiteranteil	0,0008	0,0007	0,0157	1,129	0,0001	0,0001	0,0014	0,104	0,0005	0,0007	0,0097	0,729
Auszubildendenanteil	-0,0078	0,0027	-0,0414	-2,899**	0,0023	0,0026	0,0120	0,862	-0,0136	0,0027	-0,0721	-5,091**
Bruttowertschöpfung	0,8260	0,0532	0,2298	15,530**	0,7981	0,0516	0,2215	15,450**	0,9074	0,0512	0,2525	17,709**
Frauenanteil	0,0045	0,0007	0,0100	6,886**	0,0053	0,0001	0,1231	8,639**	0,0080	0,0006	0,1881	12,879**
Konstante	-24,8937	0,8398		-29,643**	-25,2821	0,8312		-30,418**	-27,8045	0,8659		-32,109**
R <sup>2</sup> 100	23,148				23,439				25,256			
F-Wert	184,683**				187,709**				207,179**			

Fortsetzung Tabelle 11

Variable	Startwert									
	4					5				
	B	SE B	Beta	T		B	SE B	Beta	T	
Lohn	2,0183	0,0746	0,4042	27,065**		2,0365	0,0735	0,4076	27,693**	
Ausländeranteil	0,0064	0,0021	0,0419	3,049**		0,0082	0,0021	0,0525	3,877**	
Schwerbehindertenanteil	0,0139	0,0031	0,0610	4,532**		0,0220	0,0031	0,0934	7,053**	
Facharbeiteranteil	-0,0005	0,0007	-0,0105	-0,776		0,0006	0,0007	0,0115	0,852	
Auszubildendenanteil	-0,0030	0,0027	-0,0161	-1,130		-0,0040	0,0027	-0,0207	-1,492	
Bruttowertschöpfung	0,7690	0,0529	0,2137	14,528**		0,8389	0,0509	0,2335	16,490**	
Frauenanteil	0,0058	0,0006	0,1357	9,478**		0,0071	0,0006	0,1690	11,996**	
Konstante	-25,6406	0,8396		-30,539**		-26,7075	0,8297		-32,189**	
R <sup>2</sup> *100			23,388					25,244		
F-Wert			187,175**					207,049**		

\*\* :  $\alpha < 0,01$ \* :  $\alpha < 0,05$

Tabelle 12: Nachfrage nach schwerbehinderten Arbeitnehmern

Variable	Startwert											
	1				2				3			
	B	SE B	Beta	T	B	SE B	Beta	T	B	SE B	Beta	T
Konzentrationsgrad	415,4186	51,9131	0,1355	8,002**	101,1984	50,1596	0,0339	2,018*	-244,4770	50,9909	-0,0813	-4,795**
Residuum Schwerb.	6,3952	0,1252	0,6020	51,100**	6,4829	0,1228	0,6233	52,771**	6,4107	0,1240	0,6123	57,725**
Frauenanteil	0,0017	0,0025	0,0025	0,688	-0,0004	0,0023	-0,0021	-0,163	-0,0006	0,0023	-0,0034	-0,252
Beschäftigte	0,4220	0,0582	0,0979	7,257**	0,4640	0,0567	0,1104	8,186**	0,4472	0,0579	0,1058	7,721**
Facharbeiteranteil	0,0142	0,0027	0,0666	5,301**	-0,0033	0,0023	-0,0151	-1,242	-0,0050	0,0027	-0,0225	-1,855
Ausländeranteil	-0,0112	0,0085	-0,0165	-1,323	-0,0046	0,0078	-0,0074	-0,590	0,0090	0,0080	0,0142	1,118
Auszubildendenanteil	0,0428	0,0108	0,0530	3,969**	0,0152	0,0104	0,0191	1,452	0,0307	0,0106	0,0385	2,884**
Bruttowertschöpfung	-1,1473	0,2906	-0,0740	-3,948**	-0,6972	0,2809	-0,0461	-2,482*	-0,9254	0,2832	-0,0609	-3,267**
Lohn	0,2266	0,3142	0,0105	0,721	0,3522	0,3018	0,0169	1,167	0,5139	0,3170	0,0243	1,621
Anlagevermögen	1,2087	0,2558	0,1009	4,726**	-0,1232	0,2486	-0,0106	-0,496	-0,1423	0,2520	-0,0121	-0,565
Konstante	-2,5359	3,8860		-0,653	7,8095	3,8014		2,054*	9,4295	4,0465		2,330*
R <sup>2</sup> *100	40,466				40,159				39,898			
F-Wert	291,531**				287,833**				284,717**			

Fortsetzung Tabelle 12

Variable	Startwert									
	4					5				
	B	SE B	Beta	T	B	SE B	Beta	T		
Konzentrationsgrad	457,5884	51,8496	0,1466	8,825**	79,5115	50,3720	0,0264	1,578		
Residuum Schwrb.	6,4128	0,1272	0,5930	50,429***	6,4235	0,1251	0,6112	51,349**		
Frauenanteil	-0,0029	0,0024	-0,0154	-1,197	-0,0012	0,0023	-0,0069	-0,534		
Beschäftigte	0,4475	0,0592	0,1020	7,557**	0,5390	0,0584	0,1271	9,229**		
Facharbeiteranteil	0,0051	0,0028	0,0224	1,835	0,0033	0,0027	0,0151	1,224		
Ausländeranteil	0,0389	0,0082	0,0583	4,712***	0,0244	0,0083	0,0369	2,937**		
Auszubildendenanteil	0,0707	0,0108	0,0855	6,536**	0,0485	0,0107	0,0592	4,532**		
Bruttowertschöpfung	-1,5850	0,2949	-0,1003	-5,376**	-1,2084	0,2861	-0,0793	4,224**		
Lohn	0,2026	0,3168	0,0092	0,640	-0,0022	0,3118	-0,0001	-0,007		
Anlagevermögen	1,5073	0,2601	0,1236	5,796**	0,3519	0,2525	0,0299	1,393		
Konstante	-1,4309	4,0217		-0,356	10,5995	3,9687		2,671**		
R <sup>2</sup> *100		40,692				39,226				
F-Wert		294,271**				276,833**				

\*\*: $\alpha < 0,01$ \*: $\alpha < 0,05$

Tabelle 13: Nachfrage nach schwerbehinderten Arbeitnehmern für Unternehmen mit beschäftigten Schwerbehinderten

Variable	Startwert											
	1				2				3			
	B	SE B	Beta	T	B	SE B	Beta	T	B	SE B	Beta	T
Konzentrationsgrad	281,9373	52,4379	0,0920	5,377**	57,5233	50,5931	0,0193	1,137	-107,8325	51,4097	-0,0359	-2,098*
Sb. im Betrieb (Dummy)	6,2198	0,1246	0,6228	49,915**	6,3111	0,1222	0,6441	51,634**	6,2796	0,1236	0,6354	50,806**
Frauenanteil	0,0008	0,0025	0,0042	0,322	-0,0002	0,0023	-0,0012	-0,090	0,0006	0,0023	0,0037	0,274
Beschäftigte	-0,3945	0,0609	-0,0915	-6,478**	-0,3433	0,0593	-0,0817	-5,794**	-0,3631	0,0605	-0,0859	-6,007**
Facharbeiteranteil	0,0085	0,0027	0,0396	3,122**	-0,0033	0,0027	-0,0149	-1,210	-0,0056	0,0027	-0,0295	-2,075**
Ausländeranteil	-0,0094	0,0085	-0,0139	-1,096	-0,0040	0,0077	-0,0065	-0,514	0,0061	0,0081	0,0096	0,75
Auszubildendenanteil	0,0416	0,0109	0,0515	3,822**	0,0272	0,0105	0,0342	2,579**	0,0390	0,0107	0,0489	3,640**
Bruttowertschöpfung	-0,7833	0,2933	-0,0506	-2,671**	-0,5092	0,2833	-0,0336	-1,797	-0,4482	0,2853	-0,0295	-1,571
Lohn	0,4899	0,3170	0,0228	1,545	0,3106	0,3043	0,0149	1,020	0,5402	0,3191	0,0255	1,693
Anlagevermögen	0,9660	0,2581	0,0806	3,743**	-0,0072	0,2507	-0,0006	-0,029	-0,0740	0,2537	-0,0063	-0,292
Konstante	-7,3351	3,9214		-1,871	3,3612	3,8347		0,877	1,3975	4,0772		0,343
R <sup>2</sup> *100		39,414				39,137				39,074		
F-Wert		279,025**				275,780**				275,064**		

Fortsetzung Tabelle 13

Variable	4				5				
	Startwert		T		SE B		Beta		T
	B	SE B	Beta	T	B	SE B	Beta	T	
Konzentrationsgrad	321,2941	52,3888	0,1029	6,133**	21,2432	50,6478	0,0070	0,419	
Sb. im Betrieb	6,2374	0,1267	0,6126	49,214**	6,2817	0,1240	0,6359	50,645**	
Frauenanteil	-0,0038	0,0025	-0,0200	-1,536	-0,0019	0,0023	-0,0108	-0,833	
Beschäftigte	-0,3217	0,0618	-0,0733	-5,209**	-0,2575	0,0608	-0,0607	-4,236**	
Facharbeiteranteil	0,0018	0,0028	0,0079	0,642	-0,0003	0,0027	-0,0015	-0,121	
Ausländeranteil	0,0310	0,0083	0,0466	3,738**	0,0223	0,0083	0,0334	2,680**	
Auszubildendenanteil	0,0581	0,0109	0,0702	5,320**	0,0331	0,0108	0,0403	3,072**	
Bruttowertschöpfung	-1,2155	0,2976	-0,0769	-4,084**	-0,7198	0,2877	-0,0472	-2,502**	
Lohn	-0,0038	0,3197	-0,0002	-0,012	-0,4890	0,3135	-0,0231	-4,560	
Anlagevermögen	1,2361	0,2625	0,1014	4,709**	0,3414	0,2539	0,0290	1,345	
Konstante	-1,1507	4,0578		-0,284	9,3915	3,9895		2,345*	
R <sup>2</sup> *100		39,621				38,590			
F-Wert		281,447**				269,526**9			

\*\*· $\alpha < 0,01$ \*· $\alpha < 0,05$



## LITERATURVERZEICHNIS

- Aarts, E.H.L./Laarhoven, P.J.M. van (1985): Statistical Cooling: A General Approach to Combinatorial Optimization Problems, in: Philips Journal of Research, Vol. 40, No. 4, S. 193-226
- Aarts, E.H.L./Laarhoven, P.J.M. van (1987): Simulated Annealing: A Pedestrian Review of the Theory and Some Applications, in: Devijver, P.A./Kittler, J. (Hrsg.): Pattern Recognition Theory and Applications, Berlin u.a., S. 179-192
- Aarts, E.H.L./Laarhoven, P.J.M. van (1989): Simulated Annealing: An Introduction, in: Statistica Neerlandica 43, Nr. 1, S. 31-52
- Akerlof, G.A. (1982): Labor Contracts as a Partial Gift Exchange, in: Quarterly Journal of Economics, 97, S. 543-569
- Albach, H./Burg, K.-H. (1986): Empirische Untersuchungen der Firmenentwicklung, in: Krelle, W. (Hrsg.): Ökonomische Prognose-, Entscheidungs- und Gleichgewichtsmodelle, Weinheim, S. 147-167
- Albach, H./Dahremöller, A. (1986): Der Beitrag des Mittelstandes bei der Lösung von Beschäftigungsproblemen in der Bundesrepublik Deutschland, ifm-Materialien, Nr. 140
- Allegrezza, S. (1990): Das Betriebspanel in Luxemburg, in: Ertel, R./Gerlach, K./Wagner, J. (Hrsg.): Betriebsbefragungen - Bedeutung für wirtschaftswissenschaftliche Forschung und Politikberatung, NIW-Vortragsreihe, Bd. 5, S. 9-20
- Amtliche Nachrichten der Bundesanstalt für Arbeit (1988), Nr. 12, S. 1755-1757
- Anily, S./Federguen, A. (1987a): Ergodicity in Parametric Nonstationary Markov Chains: An Application on Simulated Annealing Methods, in: Operations Research, Vol. 35, Nr. 6, S. 867-874
- Anily, S./Federguen, A. (1987b): Simulated Annealing Methods with General Acceptance Probabilities, in: Journal of Applied Probabilities 24, S. 657-667
- Arminger, G./Lijphard, N./Müller, W. (1981): Die Verwendung log-linearer Modelle zur Disaggregation aggregierter Daten, in: Allgemeines Statistisches Archiv, S. 273-293
- Autorengemeinschaft (1990): Zur Konzeption der Pilotstudie für das Firmenpanel "Produzieren in Niedersachsen und Europäischer Binnenmarkt", Diskussionspapier des Fachbereichs Wirtschaftswissenschaften der Universität Hannover, Nr. 149
- Axelrod, R. (1984): The Evolution of Cooperation, New York
- Backes-Gellner, U./Böck, R./Wienecke, S. (1994): Quinter Studie der Personalpolitik in Europa (QUIPPE), Konzeption und erste Befunde, QUINT-ESSENZEN, Nr. 41
- Backhaus, K. et. al. (1989): Multivariate Analysemethoden, Berlin u.a.
- Barron, J.M./Black, D.A./Loewenstein, M.A. (1987): The Implications for Search, Training, Capital Investment, Starting Wages and Wage Growth, in: Journal of Labor Economics, Nr. 5, S. 76-89

- Barzen, D./Wahle, P. (1990): Das PIMS-Programm - was es wirklich wert ist, in: HARVARDmanager, Nr. 1, S. 100-109
- Becker, G.S. (1962): Investment in Human Capital: A Theoretical Analysis, in: Journal of Political Economy, Supplement 70, S. 9-49
- Becker, H. (1985): Eine transaktionskostentheoretische Interpretation interner Arbeitsmärkte, Berlin
- Becker, T. (1988): Zukunftsmarkt Datenbanken, in: technologie & management, Heft 2, 1988, S. 22-29
- Bellmann, L./Reinberg, A./Tessaring, M. (1992): Bildungsexpansion, Entwicklung der Qualifikationsstruktur und Einkommensverteilung, Vortrag auf der Tagung des Bildungsökonomischen Ausschusses des Vereins für Socialpolitik in Nürnberg, 8./9.10. 1992
- Birch, D. (1979): The Job Generation Process, Program on Neighbourhood and Regional Change, Cambridge/Massachusetts
- Bishop, Y.M.M./Fienberg, S.E./Holland, P.W. (1975): Discrete Multivariate Analyses: Theory and Practice, Cambridge, Mass.
- Bleymüller, J./Gehlert, G./Gühlicher, H. (1985): Statistik für Wirtschaftswissenschaftler, 4. Aufl., München
- Blien, U. (1986): Unternehmensverhalten und Arbeitsmarktstruktur. Eine Systematik und Kritik wichtiger Beiträge zur Arbeitsmarkttheorie, Nürnberg
- Blien, U./Graef, F. (1991): Entropieoptimierungsverfahren in der empirischen Wirtschaftsforschung, in: Jahrbuch für Nationalökonomie und Statistik Bd. 208/4, S. 399-413
- Bohachevsky, I.O./Johnson, M.E./Stein, M.L. (1986): Generalized Simulated Annealing for Function Optimization, in: Technometrics, Vol. 28, No 3, S. 209-217
- Bonomi, E./Lutton, J.-L. (1984): The N-City Travelling Salesman Problem: Statistical Mechanics and the Metropolis Algorithm, in: SIAM-Review. Vol. 26, No. 4, S. 551-568
- Börner, R. (1988): Jahresabschlussinformationen - ein wichtiger volkswirtschaftlicher Informationsfaktor, in: Technologie und Management 2/1988, S. 32-34
- Blossfeld, H.-P./Hannan, M.T./Schönmann, K. (1988): Erwerbsverlauf und die Entwicklung der Arbeitseinkommen bei Männern. Eine Längsschnittstudie unter Verwendung einer stochastischen Differenzialgleichung, in: Zeitschrift für Soziologie 17, Nr. 6, S. 407-423
- Blossfeld, H.-P./Hannan, M.T./Schönmann, K. (1989): Determinanten der Verdienstentwicklung auf ein und demselben Arbeitsplatz - Ein Beispiel zur Anwendung einer stochastischen Differenzialgleichung, in: Gerlach, K./Hübler, O. (Hrsg.): Effizienzlohntheorie, Individualeinkommen und Arbeitsplatzwechsel, Frankfurt a. Main, S. 301-316
- Breitung, J. (1992): Dynamische Modelle für die Paneldatenanalyse, Frankfurt a. Main
- Brown, C./Medoff, J. (1989): The Employer Size-Wage Effect, in: Journal of Political Economy, (97)5, S. 1027-1059
- Brüderl, J. (1989): Senioritätentlohnung und Effizienzlohnmodelle, in: Köhler, C./Preisendörfer, P. (Hrsg.): Betrieblicher Arbeitsmarkt im Umbruch, Frankfurt a. Main und New York, S. 75-92

- Bruse, H. (1980): Der Absatzbereich von Unternehmen - Theorie, Methoden, empirische Untersuchung, Bonn
- Büchtemann, C.F. (1989): Befristete Arbeitsverträge nach dem Beschäftigungsförderungsgesetz, Berlin
- Burgbacher, H.G./Hartwig, M./Liebau, E. (1989): Zur Empirie befristeter Arbeitsverträge, in: Sozialer Fortschritt, S. 101-104
- Buttler, F. (1987): Einführung: Einige Probleme und Wege institutionalistischer Arbeitsmarktanalyse, in: ders. et. al. (Hrsg.): Arbeitsmarkt und Beschäftigung, Frankfurt a. Main, S. 7-22
- Buzzell, R.D./Gale, B.T. (1989): Das PIMS-Programm, Strategien und Unternehmenserfolg, Wiesbaden
- Cerný, V. (1985): Thermodynamical Approach to the Traveling Salesman Problem: An Efficient Simulation Algorithm, in: Journal of Optimization Theory and Applications, Vol. 45, No. 1, S. 41-51
- Collatz, L./Wetterling, W. (1971): Optimierungsaufgaben, Berlin
- Cortese, C.F./Falk, R.F./Cohen, J.K. (1976): Further Considerations on the Methodological Analyses of Segregation Indices, in: American Sociological Review, Vol. 41, S. 630-637
- Cowherd, D.M./Luchs, R.H. (1988): Linking Organization Structures and Processes to Buisness Strategy, in: Long Range Planning, Vol. 21, No. 5, S. 47-53
- Cramer, U. (1986): Zur Stabilität von Beschäftigung, Erste Ergebnisse der IAB-Stichprobe aus der Beschäftigtenstatistik, in: MittAB, Heft 2, S. 243-256
- Cramer, U. (1987): Klein- und Mittelbetriebe: Hoffnungsträger der Beschäftigungspolitik?, in: MittAB, Heft 1, S. 15-29
- Crouchley, R./Pickles, A.R./O'Farrell, P.N. (1991): Duration of unemployment among men in Northern Ireland: regional inequality or denominational discrimination, in: Environment and Planning A, Vol. 23, S. 903-920
- Dahremöller, A. (1987): Existenzgründungsstatistik: Nutzung amtlicher Datenquellen zur Erfassung des Gründungsgeschehens, Stuttgart
- Dahremöller, A. (1988): Erstellung regionaler und betriebsgrößenbezogener Arbeitsmarktbilanzen, Stuttgart
- Deeke, A./Fischer, J. (1986): Wege der Rekrutierung auf den Arbeitsmarkt, SAMF-Arbeitspapier 1986-6
- Dialog Database Catalog 1989
- Dombois, R. (1982): Die betriebliche Normenstruktur, in: Dohse, K./Jürgens, U./Russig, H. (Hrsg.): Statussicherung im Industriebetrieb, Frankfurt a. Main/New York, S. 173-204
- Dostal, W. (1989): Arbeitsplatzwirkungen moderner Technologien, in: MittAB, S. 187-201
- Dragendorf, R./Heering, W. (1987): Beschäftigungsdauer, Effizienz und Flexibilität - Ein Beitrag zur Ökonomie des Beschäftigungsvertrages, in: Buttler, F. et. al. (Hrsg.): Arbeitsmarkt und Beschäftigung, Frankfurt a. Main u. New York, S. 121-156

- Duda, H. (1987): Macht oder Effizienz? Eine ökonomische Theorie der Arbeitsbeziehungen im modernen Unternehmen, Frankfurt a. Main u.a.
- Eglese, R.W. (1990): Simulated Annealing: A tool for Operational Research, in: *European Journal of Operational Research* 46, S. 271-281
- Engel, B./Krause, P./Wagner, G. (1991): Datenbankorganisatorische Probleme und Grundlagen des NIFA-Panels, Ergebnisse eines Workshops, Arbeitspapier des SFB 187 der Ruhr-Universität Bochum
- Erhard, U. (1983): Zur Homogenität der Unternehmen und der Wirtschaftsgruppen im Verarbeitenden Gewerbe, in: *Allgemeines Statistisches Archiv*, S. 397-375
- Evans, D.S./Leighton, L.S. (1988): Why Do Smaller Firms Pay Less? in: *The Journal of Human Resources*, (24)2, S. 299-318
- Faigle, U./Schrader, R. (1988): Simulated Annealing - Eine Fallstudie, in: *Angewandte Informatik*, Heft 6, S. 259-263
- Falk, R.F./Cortese, C.F./Cohen, J. (1978): Utilizing Standardized Indices of Residential Segregation, in: *Social Forces*, Vol. 57, S. 713-716
- Fitzroy, F.R./Kraft, K. (1987): Formen der Arbeitnehmer-Arbeitgeberkooperation und ihre Auswirkung auf die Unternehmensleistung und Entlohnung, in: Dies. (Hrsg.): *Mitarbeiter-Beteiligung und Mitbestimmung im Unternehmen*, Berlin/New York, S. 173-196
- Fleury, G. (1991): Simulated Annealing and Scheduling in Manufacturing Systems, Vortragsmanuskript anlässlich eines Vortrags an der Universität Trier
- Flimm, C./Saurwein, R. (1991): Das NIFA-Panel: Gegenstand und Aufbau, Arbeitspapier des SFB 187 der Ruhr-Universität Bochum
- Foster, A. (1984): Databases and databanks, in: Fletcher, J. (ed.): *Information Sources in Economics*, 2nd ed., S. 73-95
- Foster, A. (1984a): Online Services and computer data banks for business, in: Vernon, K.D.C. (ed.): *Information Sources in Management and Business*, 2nd ed., London u. a., S. 160-188
- Freriks, R. (1991): Das NIFA-Panel: Theoretische Grundlagen und Implikationen, Arbeitspapier des SFB 187 der Ruhr-Universität Bochum
- Frick, B. (1991): Interne Arbeitsmärkte und betriebliche Schwerbehindertenbeschäftigung: Theoretische Analysen und empirische Befunde, in: *Zeitschrift für Personalforschung*, 5. Jg., Heft 4, S. 309-313
- Frick, B. (1992): Interne Arbeitsmärkte und betriebliche Schwerbehindertenbeschäftigung: theoretische Analysen und empirische Befunde, Frankfurt a. Main/New York
- Frick, B./Sadowski, D. (1988): Zur Beschäftigungssituation Schwerbehinderter. Eine empirische Analyse anhand des Sozio-ökonomischen Panels, SAM-Diskussionsbeitrag Nr. 17, Trier
- Friedrichs, J. (1985): *Methoden empirischer Sozialforschung*, 13. Aufl., Opladen
- Fritsch, M. (1990a): *Arbeitsplatzentwicklung in Industriebetrieben*, Berlin/New York
- Fritsch, M. (1990b): Wie heterogen sind Branchenaggregate?, in: *Allgemeines Statistisches Archiv* 74, S. 397-412

- Fritsch, M./Hull, C.J. (1987): Empirische Befunde zur Arbeitsplatzdynamik in großen und kleinen Unternehmen in der Bundesrepublik Deutschland - Eine Zwischenbilanz -, in: Dies. (Hrsg.): Arbeitsplatzdynamik und Regionalentwicklung. Beiträge zur beschäftigungspolitischen Bedeutung von Klein- und Großunternehmen, Berlin, S. 149-172
- Fuchs, W. et al. (1978): Lexikon zur Soziologie, 2. Aufl., Opladen
- Geisen, B. (1976): Das Datenmaterial der Bonner Stichprobe, Diskussionspapier, Bonn
- Gerlach, K./Hübler, O. (Hrsg. 1989): Effizienzlohntheorie, Individualeinkommen und Arbeitsplatzwechsel, Frankfurt - New York
- Gerlach, K./Schasse, U. (1988): On-the-Job Training Differences by Sex and Firm Size, SAMF-Arbeitspapier 1988-5, Paderborn
- Gerlach, K./Schmidt, E.M. (1989): Unternehmensgröße und Entlohnung, in: Mitteilungen aus der Arbeitsmarkt- und Berufsforschung, Heft 3, S. 355-373
- Gerlach, K. u. a. (1990): Zu konzeptionellen und methodischen Aspekten eines Betriebspanels, Projektbericht, erstellt im Auftrag der Bundesanstalt für Arbeit, Hannover
- Gerlach, K./Wagner, J. (1992): Die Beschäftigungsdynamik im Bergbau und Verarbeitenden Gewerbe in Niedersachsen: Eine Komponentenanalyse für den Zeitraum 1978 bis 1990, in: Statistische Monatshefte Niedersachsen, Heft 1, S. 5-10
- Gibbs, J.P. (1965): Occupational Differentiation of Negroes and Whites in the United States, in: Social Forces, Vol. 44, S. 159-165
- Görzig, B. (1991): Zeitreihen des Anlagevermögens in Industriezweigen nach Beschäftigtengrößenklassen, in: Deutsches Institut für Wirtschaftsforschung, Beiträge zur Strukturforchung, Heft 123
- Gottfredson, L./White, P. (1981): Interorganizational Agreements, in: Nyström, P.C./Starbuck, W. (Hrsg.): Handbook of Organizational Design, Vol. 1, S. 471-486
- Graef, F./Blien, U. (1989): ENTROP: Ein allgemein einsetzbares Verfahren zur Gewichtung von Stichproben, zur Disaggregation von Daten und zur Ermittlung von Tabellen aus heterogenen Informationen, in: Allgemeines Statistisches Archiv 73, S. 122-142
- Gray, R.M. (1990): Entropy and Information Theory, New York u.a.
- Griliches, Z. (1986): Economic Data Issues, in: Griliches, Z./Intriligator, M.D. (ed.): Handbook of Econometrics, Vol. 3, Amsterdam, S. 1465-1514
- Groshen, E.L. (1991): Sources of Intra-Industry Wage Dispersion: How much do Employers matter?, in: The Quarterly Journal of Economics, August 1991, S. 869-884
- Grüner, H. (1989): Karrieremuster und soziale Differenzierung im betrieblichen Arbeitsmarkt, in: Köhler, C./Preisendörfer, P. (Hrsg.): Betrieblicher Arbeitsmarkt im Umbruch, Frankfurt a. Main und New York, S. 93-116
- Gyárfás, G. (1988): Konstruktion einer synthetischen Mikrodatenbasis für die Simulation der Einkommensteuer, Arbeitspapier der Gesellschaft für Mathematik und Datenverarbeitung Nr. 325

- Gyárás, G. (1990): Ein Simulationsmodell der Einkommensbesteuerung auf der Grundlage synthetischer Mikrodaten, Dissertation, Universität Bonn
- Gyárfás, G./Quinke, H. (1991): Ein Verfahren zur Konstruktion synthetischer Mikrodaten aus aggregierten Strukturinformationen, Diskussionspapier Nr. A-346 der Rheinischen Friedrich-Wilhelms-Universität Bonn
- Gyárfás, G./Quinke, H. (1993): Ein Verfahren zur Konstruktion synthetischer Mikrodaten aus aggregierten Daten, in: Allgemeines Statistisches Archiv 77, S. 149-165
- Hajek, B. (1988): Cooling Schedules for Optimal Annealing, in: Mathematics of Operations Research, Vol. 13, No. 2, S. 311-329
- Hamermesh, D.S. (1988): Data Difficulties in Labor Economics, NBER Working Paper #2622
- Hamermesh, D.S. (1993): Labor Demand, Princeton, New Jersey
- Handl, J. (1984): Chancengleichheit und Segregation: Ein Vorschlag zur Messung ungleicher Chancenstrukturen und ihrer zeitlichen Entwicklung, in: Zeitschrift für Soziologie, Vol. 13, S. 328-345
- Hanefeld, U. (1984): Das Sozio-ökonomische Panel - Eine Längsschnittstudie für die Bundesrepublik Deutschland, in: Vierteljahreshefte zur Wirtschaftsforschung, Heft 4, S. 391-406
- Hanefeld, U. (1985): Das Sozio-ökonomische Panel - Konzeption und ausgewählte erhebungsmethodische Ergebnisse, in: Allgemeines Statistisches Archiv, Vol. 69, Heft 4, S. 399-410
- Hartmann, H. (1979): Qualität und Quantität des Arbeitskräfteangebots in ihrer Auswirkung auf unternehmerische Entscheidungen, in: Oppenländer, K.-H. (Hrsg.): Unternehmerischer Handlungsspielraum in der aktuellen wirtschafts- und sozialpolitischen Situation, Berlin, S. 149-172
- Hauptmanns, P. (1990): Das NIFA-Panel: Untersuchungsanlage und methodische Konzeption, Arbeitspapier des SFB 187 der Ruhr-Universität Bochum
- Hauptmanns, P./Saurwein, R.G. (1992): Probleme bei schriftlichen Betriebsbefragungen: Ein Erfahrungsbericht zur ersten Welle des NIFA-Panels, in: ZUMA-Nachrichten, 31. Jg., 16. November 1992, S. 70-86
- Heller, W.D./Lindenberg, H./Nuske, M./Schriever, K.H. (1978): Stochastische Systeme, Berlin/NewYork
- Hengartner, W./Theodorescu, R. (1978): Einführung in die Monte-Carlo-Methode, München u. Wien
- Hohn, H.-W. (1989): Ausgrenzung und Stabilisierung der Massenarbeitslosigkeit: Gelöste und ungelöste Probleme im "dualen" System der Interessenvertretung, in: Peters, W. (Hrsg.): Massenarbeitslosigkeit und Politik: Reaktionsweisen und Strategieoptionen in verschiedenen Politikarenen, SAMF-Arbeitspapier 1989-1
- Hohwü, J. (1990): Interregionale Paneluntersuchungen von Betrieben in Süd-Dänemark, in: Ertel, R./Gerlach, K./Wagner, J. (Hrsg.): Betriebsbefragungen - Bedeutung für wirtschaftswissenschaftliche Forschung und Politikberatung, NIW-Vortragsreihe, Bd. 5, S. 21-32

- Hotchkiss, J.M. (1990): Compensation Policy and Firm Performance: An Associated Bibliography of Machine-Readable Data Files, in: *Industrial Labor Relations Review*, Vol. 43, S. 274-289
- Hübler, O. (1990): Lineare Paneldatenmodelle mit alternativer Größenstruktur, in: Nakhaeizadeh, G./Vollmer, K.-H. (Hrsg.): *Neuere Entwicklungen in der angewandten Ökonometrie*, Heidelberg, S. 65-99
- Hübler, O./Gerlach, K. (1989): Berufliche Lohndifferentiale, in: Gerlach, K./Hübler, O. (Hrsg.): *Effizienzlohntheorie, Individualeinkommen und Arbeitsplatzwechsel*, Frankfurt - New York, S. 149-183
- Huinink, J. (1989): Mehrebenensystem-Modelle in den Sozialwissenschaften, Wiesbaden
- Institut für Arbeitsmarkt- und Berufsforschung, Projektgruppe Betriebspanel (1992): *IAB Betriebspanel: Projektstand, Testergebnisse und konzeptionelle Folgerungen*, Nürnberg
- Jaglom, A.M./Jaglom, I.M. (1984): *Wahrscheinlichkeit und Information*, 4. Aufl., Thun u. Frankfurt a. M.
- Johnson, D.S./Aragon, C.R./McGeoch, L.A./Schevon, C. (1989): Optimization by Simulated Annealing: An Experimental Evaluation; Part I, Graph Partitioning, in: *Operations Research*, Vol. 37, No. 6, S. 865-892
- Johnson, D.S./Aragon, C.R./McGeoch, L.A./Schevon, C. (1991): Optimization by Simulated Annealing: An Experimental Evaluation; Part II, Graph Coloring and Number Partitioning, in: *Operations Research*, Vol. 39, No. 3, S. 378-406
- Jörnsten, K.O./Jarsson, T./Lundgren, J.T./Migdalas, A. (1990): An entropy model with variable target, in: *Environment and Planning A*, Vol. 22, S. 493-506
- Kirkpatrick, S./Gelatt C.D. jr./Vecchi, M.P. (1983): Optimization by Simulated Annealing, in: *Science*, Vol. 220, No. 4598, S. 671-680
- Köhler, C./Preisendörfer, P. (Hrsg., 1989): *Betrieblicher Arbeitsmarkt im Umbruch*, Frankfurt a. Main und New York
- Köhler, C./Schultz-Wild, R. (1989): Der gemeinsame Forschungsgegenstand "Südwerk": Struktur und Entwicklung eines betrieblichen Arbeitsmarktes, in: Köhler, C./Preisendörfer, P. (Hrsg.): *Betrieblicher Arbeitsmarkt im Umbruch*, Frankfurt/m. u. a., S. 11-30
- Köhler, C./Sengenberger, W. (1983): *Konjunktur und Personalanpassung*, Frankfurt a. Main/New York
- Kohli, M./Wolf, J. (1987): Altersgrenzen im Schnittpunkt von betrieblichen Interessen und individueller Lebensplanung. Das Beispiel des Vorruhestands, in: *Soziale Welt*, 38, S. 92-109
- Krüger, S. (1975): *Simulation. Grundlagen, Techniken, Anwendungen*, New York
- Kruse, D. (1991): Supervision, Working Conditions, and the Employer Size-Wage Effect, in: *The Journal of Industrial Relations*, (33)3, S. 229-249

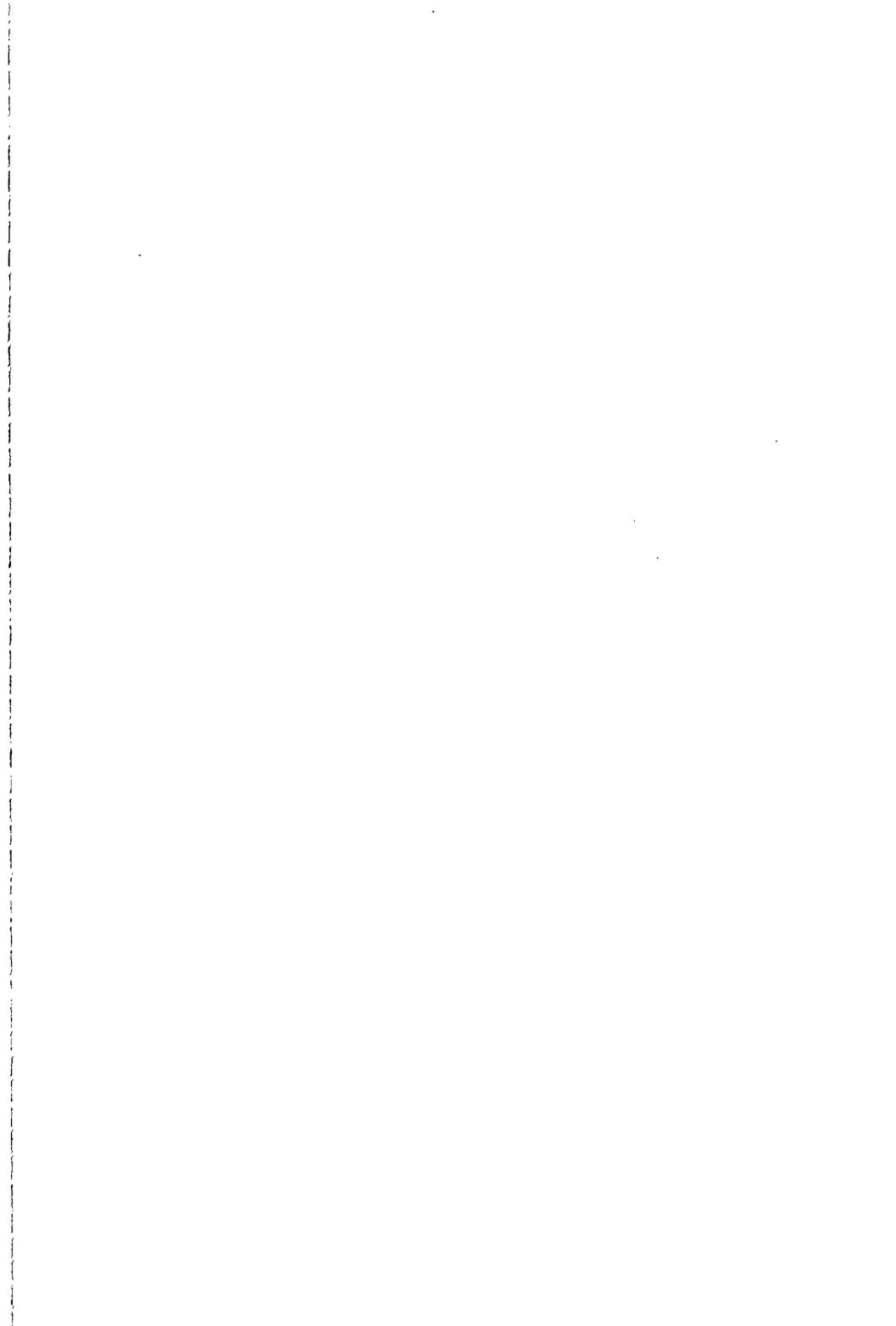
- Kühl, J. (1995): Betriebsentwicklung und Beschäftigung. Erste Ergebnisse des IAB-Betriebspanels aus der Befragungswelle 1993, in: Semlinger, K./Frick, B. (Hrsg.): Betriebliche Modernisierung in personeller Erneuerung, Berlin, S. 25-38
- Kuhn, H. (1990): Einlastungsplanung von flexiblen Fertigungssystemen, Heidelberg
- Kuhn, H. (1992): Heuristische Suchverfahren mit simulierter Abkühlung, in: Wirtschaftswissenschaftliches Studium, Heft 4, S. 387-391
- Laarhoven, P.J.M. van (1988): Theoretical and Computational Aspects of Simulated Annealing, Dissertation, Universität Rotterdam
- Laarhoven, P.J.M. van/Aarts, E.H.L. (1987): Simulated Annealing: Theory and Applications, Dordrecht
- Leontief, W. (1982): "Academic Economics", Letter to the Editors, in: Science 217, S. 104-107
- Levine, R.D./Tribus, M. (ed., 1978): The Maximum Entropy Formalism, Cambridge u. a.
- Lundy, M./Mees, A. (1985): Convergence of an Annealing Algorithm, in: Mathematical Programming, Vol. 34, S. 111-124
- Lutz, B. (1987): Arbeitsmarktstruktur und betriebliche Arbeitskräftestrategie, Frankfurt a. Main
- Marshall, A. (1961): Principles of Economics, 9. Aufl., London
- Massey, D.S. (1978): On the Measurement of Segregation as a Random Variable, in: American Sociological Review, Vol. 43, S. 587-590
- Matzner, E./Schettkat, R./Wagner, M. (1988): Beschäftigungsrisiko Innovation? Arbeitsplatzwirkungen moderner Technologien. Befunde aus der Meta-Studie, Berlin
- Mendius, H.-G./Sengenberger, W. (1976): Konjunkturschwankungen und betriebliche Politik. Zur Entstehung und Verfestigung von Arbeitsmarktsegmentation, in: Dies. et al. (Hrsg.): Betrieb - Arbeitsmarkt - Qualifikation, Frankfurt a. Main, S. 15-81
- Merz, J. (1983): Die konsistente Hochrechnung von Mikrodaten nach dem Prinzip des minimalen Informationsverlustes, in: Allgemeines Statistisches Archiv, S. 342-366
- Metropolis, N. et. al. (1953): Equation of State Calculations by Fast Computing Machines, in: Journal of Chemical Physics 21, S. 1087-1092
- Meyer, J./Greif, H.-H. (1990): Strategische Informationssysteme. PIMS - Das Instrument zur Strategischen Kursbestimmung im Zeitalter der Diskontinuität, Sonderdruck aus: Strategisches Management in den 90er Jahren
- Mincer, J. (1962): On-the-Job Training: Costs, Returns and Some Implications, in: Journal of Political Economy, Supplement, Vol. 70, S. 50-79
- Mincer, J. (1974): Schooling, Experience and Earning, New York
- Mitra, D./Romeo, F./Sangiovanni-Vincentelli, A. (1986): Convergence and Finite-Time Behavior of Simulated Annealing, in: Advances in Applied Probability, Vol. 18, S. 747-771

- Monk, T.J./Landis, K.M./Monk, S.S. (1988): The Dow Jones-Irwin Banker's Guide to Online Databases, Homewood, Illinois
- Naschold, F./Ragner, G./Rosenow, J. (1989): Betriebe und Staat im altersstrukturellen Wandel - eine Projektskizze, unv. Ms., Berlin
- Niggemann, H. (1993): Die Anpassung von Mikrodaten an Restriktionen mit dem statischen Mikrosimulationsmodell MICSIM: Ein Konzept zur Analyse von Ausfallprozessen im NIFA-Panel, Arbeitspapier Z2-3/93 des SFB 187 der Ruhr Universität Bochum
- Nissan, E./Caveny, R. (1988): Relativ Concentration of Sales and Assets in American Buisness, in: Southern Economic Journal, Vol 45, Nr. 4, S. 928-933
- Oi, W.Y. (1962): Labor as a Quasi-Fixed Factor, in: Journal of Political Economy, Vol. 70, S. 538-555
- Oppenländer, K.-H./Poser, G. (Hrsg. 1989): Handbuch der Ifo-Umfragen, Berlin/München
- Orcutt, G.H. (1990): The Microanalytic Approach for Modeling National Economies, in: Journal of Economic Behavior and Organizations, Vol. 14, S. 29-41
- Orcutt, G.H./Watts, H./Edwards, J. (1968): Data Aggregation and Information Loss, in: American Economic Review, Vol. 58, S. 773-787
- Ostendorf, B./Schmid, J. (1994): Der Einfluß der Betriebsgröße. Analysen und Interpretationen der zweiten Welle des NIFA-Panels, Arbeitspapier Z2-1/94 des SFB 187 der Ruhr Universität Bochum
- Paaß, G. (1988): Stochastic Generation of a Synthetic Sample from Marginal Information, Arbeitspapier der Gesellschaft für Mathematik und Datenverarbeitung Nr. 308
- Penzkofer, H. (1990): Standardbefragungen des Ifo-Instituts für Wirtschaftsforschung, in: Ertel, R./Gerlach, K./Wagner, J. (Hrsg.): Betriebsbefragungen. Bedeutung für wirtschaftswissenschaftliche Forschung und Politikberatung, NIW-Vortragsreihe, Band 5, Hannover, S. 59-74
- Pietsch, P. (1984): Statistische Probleme der Ermittlung von industriellen Arbeitsproduktivitäten, Frankfurt a.Main
- Projektgruppe Betriebspanel (1991): Das IAB-Betriebspanel - Ansatz und Aufbau, in: MittAB, Heft 3, S. 514-530
- Projektgruppe Betriebspanel (1994): IAB-Betriebspanel - Ergebnisse der ersten Welle 1993, in: MittAB, Heft 1, S. 20-32
- Psacharopoulos, G. (1985): Returns to Education: A Further International Update and Implications, in: Journal of Human Resources, Vol. 20, S. 583-597
- Ritz, H.-G. (1984): Arbeitsmarktlage und Novellierung des Schwerbehindertengesetzes, Arbeitspapier Nr. 22 des Forschungsschwerpunktes Reproduktionsrisiken, soziale Bewegung und Sozialpolitik, Universität Bremen, Bremen
- Rogalski, W. (1994): Dokumentation der NIFA-Datenbank, Dokumentations- und Informationspapier Z2-4/94 des SFB 187 der Ruhr Universität Bochum

- Rothe, G. (1990): Wie (un)wichtig sind Gewichtungen? Eine Untersuchung am ALLBUS 1986, in: ZUMA-Nachrichten 1990, S. 31-55
- Sadowski, D. (1989): Währt ehrlich am längsten? Personalpolitik zwischen Arbeitsrecht und Unternehmenskultur, in: Budäus et al.: Betriebswirtschaftslehre und Theorie der Verfügungsrechte, Frankfurt a. Main, S. 220-238
- Sadowski, D./Frick, B. (1992): Die Beschäftigung Schwerbehinderter: Betriebswirtschaftliche Analysen und politische Empfehlungen, Idstein
- Sadowski, D./Böck, R./Brühl, N./Frick, B. (1992): Die Wirkungsweise des Schwerbehindertengesetzes - Vollzugsdefizite und Verbesserungsvorschläge -, Forschungsbericht des Bundesministeriums für Arbeit und Sozialordnung, Nr. 230
- Sadowski, D./Stengelhofen, T. (1989): Betriebliche Theorie der Nichtlohn-Arbeitskosten am Beispiel der Fluktuation der Arbeitnehmer, in: Emmerich, K. (Hrsg.): Einzel- und gesamtwirtschaftliche Aspekte des Lohnes, BeitrAB 128, Nürnberg, S. 103-116
- Salop, S. (1979): A Model of the Natural Rate of Unemployment, in: American Economic Review, 69, S. 117-125
- Samuelson, P.A./Nordhaus, W.E. (1985): Economics, 12th ed., New York u.a.
- Schasse, U. (1992): Überlebenschancen und Beschäftigungswirkungen neugegründeter Industriebetriebe in Niedersachsen, in: Ertel, R./Wagner, J. (Hrsg.): Produzieren in Niedersachsen. Empirische Untersuchungen mit Betriebsdaten, NIW-Vortragsreihe, Band 6, Hannover, S. 31-60
- Scherer, F.M. (1986): Stand und Perspektiven der Industrieökonomik, in: Bombach, G./Gahlen, B./Ott, A.E. (Hg.): Industrieökonomik: Theorie und Empirie, Tübingen, S. 3-19
- Schmid, G. (1987): Grundzüge einer Theorie institutionalisierter Arbeitsbeziehungen, in: Buttler et al. (Hrsg.): Arbeitsmarkt und Beschäftigung, Frankfurt a. Main u. New York, S. 24-63
- Schmidt, H.M. (1988): Freiwillig ja - zwangsweise nein, in: technologie & management, Heft 2, 1988, S. 34-35
- Schneeweis, H. (1990): Ökonometrie, 4. Aufl., Heidelberg
- Scholl, W. (1987): Betriebliche Strategien zur Befriedigung von Arbeitnehmerbedürfnissen - Voraussetzungen und Konsequenzen, in: Fitzroy, F.R./Kraft, K. (Hrsg.): Mitarbeiter-Beteiligung und Mitbestimmung im Unternehmen, Berlin/New York, S. 131-158
- Schumacher, S./Kistner, K.-P. (1989): Unternehmensstatistiken, Discussionpaper Nr. 200 der Fakultät für Wirtschaftswissenschaften der Universität Bielefeld
- Sengenberger, W. (1987): Struktur und Funktionsweise von Arbeitsmärkten. Die Bundesrepublik Deutschland im internationalen Vergleich, Frankfurt a. Main
- Shannon, C.E. (1948): A Mathematical Theory of Communication, in: Bell System Technical Journal 27, S. 379-423, S. 623-656
- Shapiro, C./Stiglitz, J.E. (1984): Equilibrium Unemployment as a Worker Discipline Device, in: American Economic Review, 74, S. 433-444
- Shephard, R.W. (1970): Theory of Cost and Production Functions, Princeton

- Shore, J.E./Johnson, R.W. (1980): Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy, in: IEEE Transactions on Information Theory, Vol. IT-26, No. 1, S: 26-37
- Snickars, F./Weibull, J.W. (1977): A Minimum Information Principle, in: Regional Science and Urban Economics 7, S. 137-168
- Soete, G. de (1988): On the Use of Simulated Annealing for Combinatorial Data Analyses, in: Gaul, W./Schader, M. (Hrsg.): Data, Expert Knowledge and Decisions: An Interdisciplinary Approach with Emphasis on Marketing Applications, New York u.a., S. 329-340
- Sofianopoulou, S. (1992): Simulated Annealing applied to the process allocation problem, in: European Journal of Operational Research, Vol. 60, S. 327-334
- Stafford, F. (1986): Forestalling The Demise Of Empirical Economics: The Role Of Microdata In Labor Economics Research, in: Ashenfelter, O./Layard, R. (ed.): Handbook of Labor Economics, Volume I, Amsterdam, S. 387-423
- Stahl, C. (1991): Das Mannheimer Unternehmenspanel: Konzept und Entwicklung, unveröffentlichtes Manuskript
- Stäglin, R. (1970): Estimating an Input-Output Matrix with Incomplete Information, in: Konjunkturpolitik 10, S. 145-153
- Statistisches Bundesamt 1989, Fachserie 2, Heft 1
- Statistisches Bundesamt 1990, Fachserie 2, Heft 11
- Statistisches Bundesamt 1991, Fachserie 2, Heft 10
- Stiglitz, J.E. (1987): The Causes and Consequences of the Dependence of Quality on Price, Journal of Economic Literature, 25, S. 1-48
- Stinchcombe, A.L. (1965): Social Structure and Organizations, in: March, J.G. (Hrsg.): Handbook of Organizations, Chicago, S. 142-193
- Strigel, W.H. (1989): Synopse der Ifo-Umfragen, in: Oppenländer, K.-H./Poser, G. (Hrsg.): Handbuch der Ifo-Umfragen, Berlin/München, S. 334-343
- Taeuber, K.F./Taeuber, A.F. (1976): A Practitioner's Perspective on the Index of Dissimilarity, in: American Sociological Review, Vol. 41, S. 884-889
- Thaler, R.H. (1989): Anomalies: Interindustry Wage Differentials, in: Journal of Economic Perspectives, 3, S. 181-193
- Theil, H./Fiebig, D.G. (1984): Exploiting Continuity, Maximum Entropy Estimation of Continuous Distributions, Cambridge, Massachusetts
- Thurow, L.C. (1975): Generating Inequality. Mechanisms of of Distribution in the US Economy, New York
- Tudor, J. (1989): Macmillan directory of business information sources, 2nd ed., Brunel Road u.a.
- Unger, R. (1986): Messung und Analyse der totalen Faktorproduktivität für 28 Sektoren der Bundesrepublik Deutschland 1960 bis 1981, Frankfurt a. Main u. a.
- Varian, H.L. (1981): Mikroökonomie, München u.a.
- Varian, H.L. (1990): Intermediate Microeconomics, A Modern Approach, 2nd ed., New York und London

- Vassiliou, P.-G., (1984): Entropy as a Measure of the Experience Distribution in a Manpower System, in: Journal of Operational Research, Vol. 35, Nr. 11, S. 1021-1025
- Wauschkunn, U. (1982): Anpassung von Stichproben und n-dimensionalen Tabellen an Randbedingungen. Bewertung und Vergleich existierender Verfahren, München
- Weber, , H.K. (1983): Rentabilität, Produktivität, Liquidität der Unternehmung, Stuttgart
- Windolf, P./Hohn, H.-W. (1984): Arbeitsmarktchancen in der Krise, Frankfurt a. Main
- Winship, C. (1977): A Revaluation of Indexes of Residential Segregation, in: Social Forces, Vol. 55, S. 1058-1066
- Witte, T. (1973): Simulationstheorie und ihre Anwendung auf betriebliche Systeme, Wiesbaden
- Wolf, J./Kohli, M. (1988): Neue Altersgrenzen des Arbeitslebens. Betriebliche Interessen und biographische Perspektiven, in: Rosenmayr, L./Kolland, F. (Hrsg.): Arbeit-Freizeit-Lebenszeit. Neue Übergänge im Lebenszyklus, Opladen, S. 183-206
- Yellen, J.L. (1984): Efficiency Wage Models of Unemployment, in: American Economic Review, Papers and Proceedings 74, S. 200-205
- Zellner, A./Highfield, R.A. (1988): Calculation of Maximum Entropy Distributions and Approximation of Marginalposterior Distributions, in: Journal of Econometrics 37, S. 195-209



# Organisationsökonomie humaner Dienstleistungen

herausgegeben von Dieter Sadowski

*Hans Helmut Kehr*

**Leistungsorientierter Krankenhaus-Betriebsvergleich. Entwicklung eines Informations- und Kontrollsystems zur Vergleichbarkeit der Krankenhäuser hinsichtlich ihrer Wirtschaftlichkeit**

Band 2, ISBN 3-87988-129-4, Rainer Hampp Verlag, München u. Mering 1995, 198 S., DM 39.80

*Karl-Heinrich Bruckschen*

**Der Einfluß leitender Ärzte auf den technologischen Fortschritt im Krankenhaus. Ein Modell der Beschaffung medizinisch-technischer Geräte (MTG)**

Band 2, ISBN 3-87988-130-8, Rainer Hampp Verlag, München u. Mering 1995, 195 S., DM 39.80

*Norbert Brühl*

**Personalabbau und Altersstruktur. Eine betriebswirtschaftliche Analyse**

Band 3, ISBN 3-87988-246-0, Rainer Hampp Verlag, München u. Mering 1997, 270 S., DM 52.80

*Rainer Krütten*

**Synthetische Unternehmensstichproben. Ein alternativer Weg zur empirischen Analyse der Arbeitsnachfrage und deren Einfluß auf die Wertschöpfung im Unternehmen?**

Band 4, ISBN 3-87988-255-X, Rainer Hampp Verlag, München u. Mering 1997, 220 S., DM 46.80

*Hans Helmut Kehr: Leistungsorientierter Krankenhaus-Betriebsvergleich. Entwicklung eines Informations- und Kontrollsystems zur Vergleichbarkeit der Krankenhäuser hinsichtlich ihrer Wirtschaftlichkeit*

Band 2, ISBN 3-87988-129-4, Rainer Hampp Verlag, München u. Mering 1995, 198 S., DM 39.80

Durch die Neuordnung der Krankenhausfinanzierung in den letzten Jahren gewinnen Fragen der Wirtschaftlichkeit und Vergleichbarkeit von Krankenhäusern an Aktualität und Wichtigkeit. Der Gesetz- und Verordnungsgeber hat erkannt, daß ein höheres Maß an Vergütungsgerechtigkeit und -objektivität wünschenswert ist, und hat festgelegt, daß bei der Ermittlung der Pflegesätze für ein Krankenhaus auch ein Vergleich mit anderen Krankenhäusern zu erfolgen hat.

Durch den Wegfall der garantierten Selbstkostendeckung sind die Krankenhäuser zu einer wirtschaftlichen Betriebsführung gezwungen. Notwendig ist deshalb ein Controlling im Krankenhaus, das leistungsbezogene Vergleiche mit anderen Krankenhäusern ermöglicht und Verlustrisiken vermeiden hilft.

Die Veröffentlichung setzt bei der Entwicklung eines Informations- und Kontrollsystems direkt bei den Leistungen an und liefert — im Gegensatz zu allen bisherigen Verfahren — eine maßgebliche Erweiterung und Verbesserung der Vergleichbarkeit von Krankenhäusern. Ein Output-nahes Verfahren der Leistungsmessung, welches das Spektrum der behandelten Krankheiten berücksichtigt, wird entwickelt. Tabellen und Grafiken verdeutlichen konkrete Einsatz- und Anwendungsmöglichkeiten des Verfahrens.

*Norbert Brüh*

**Personalabbau und Altersstruktur. Eine betriebswirtschaftliche Analyse**

Band 3, ISBN 3-87988-246-0, Rainer Hampp Verlag, München u. Mering 1997, 198 S., DM 52.80

Von konjunkturell bedingtem Beschäftigungsabbau sind die verschiedenen Altersgruppen am Arbeitsmarkt nicht gleichmäßig, sondern die jüngsten wie die ältesten Arbeitskräfte überproportional betroffen. Zu den von den Unternehmen am häufigsten angewendeten Anpassungsinstrumenten zählen einerseits Rekrutierungs- und Übernahmestopps, die Kürzung von Ausbildungsbudgets und die Entlassung jüngerer Arbeitnehmer, andererseits kosten-trächtige Frühverrentungsprogramme.

Die Zwangsläufigkeit dieser altersselektiven Personalanpassungen resultiert nicht zuletzt aus dem empirischen Faktum meist stabiler Beschäftigungsverhältnisse. Entlang der Senioritäts- und Altersdimension der Belegschaften streuen sowohl die Entlassungskosten als auch das Verhältnis von Produktivität und Lohn als wesentliche Einflußgrößen der Personalabbauentscheidung. Um diese Zusammenhänge zu untermauern und zu quantifizieren, werden im Zuge dieser Arbeit unterschiedliche arbeitsökonomische Theorien herangezogen, die auf implizite Vereinbarungen im Rahmen von dauerhaft angelegten Arbeitsverträgen abstellen. Darüber hinaus werden die institutionellen Rahmenbedingungen – insbesondere einschlägige Regelungen des Arbeits- und Sozialrecht – analysiert. Ein investitionstheoretisches Entscheidungsmodell verknüpft die verschiedenen Einflußfaktoren; seine simulative Anwendung führt zu Ergebnissen, die mit dem Ausgangsphänomen und den Befunden der Sekundäranalyse einer Betriebsbefragung in Einklang stehen.

## Weitere Veröffentlichungen im Rainer Hampp Verlag

### **Kerstin Pull: Übertarifliche Entlohnung und freiwillige betriebliche Leistungen. Personalpolitische Selbstregulierung als implizite Verhandlung**

*International vergleichende Schriften zur Personalökonomie und Arbeitspolitik*, hrsg. von Rolf Birk und Dieter Sadowski, Band 2, ISBN 3-87988-176-6, Rainer Hampp Verlag, München und Mering 1996, 247 S., DM 46.80

Unternehmen leisten mehr als das, wozu sie rechtlich verpflichtet sind: Sie entlohnen über Tarif, sie gewähren freiwillige betriebliche Sozialleistungen, sie räumen ihren Arbeitnehmern über das Betriebsverfassungsgesetz hinaus Mitspracherechte ein - und das in einer Zeit, in der die Belastung der Unternehmen im Bereich der Personalkosten durch gesetzliche und tarifvertragliche Mindeststandards vielfach bereits als unzumutbar angesehen wird. Was ist das Kalkül, das hinter der betrieblichen Entscheidung zu einem Überbieten rechtlicher Standards steht, welches sind die Einflußfaktoren der freiwilligen Mehrleistung von Unternehmen? Warum betreiben gerade die großen Unternehmen eine eher großzügige Personalpolitik, warum unterscheiden sich Branchen, warum Regionen? Warum werden Führungskräfte bei der Leistungszuteilung so offensichtlich bevorzugt?

Auf all diese Fragen läßt sich eine Antwort geben, wenn die Entscheidung der Unternehmen für ein bestimmtes Leistungsniveau als Ergebnis einer „impliziten Verhandlung“ zwischen Unternehmensleitung und Arbeitnehmern modelliert wird. Dies hängt dann im Ergebnis sowohl von unternehmensspezifischen Faktoren wie der Arbeitsintensität und dem Ausmaß spezifischer Investitionen ab als auch von den Konkurrenzbedingungen an Absatz- und Beschaffungsmärkten sowie von rechtlichen Mindeststandards. Alle Implikationen des Modells „impliziter Verhandlung“ sind mit den stilisierten Fakten des Arbeitsmarktes kompatibel - und das, obwohl der Betriebsrat als Institution aus drücklich vernachlässigt wird.

### **Matthias Kräkel: Ökonomische Analyse der betrieblichen Karrierepolitik**

*International vergleichende Schriften zur Personalökonomie und Arbeitspolitik*, hrsg. von Rolf Birk und Dieter Sadowski, Band 8, ISBN 3-87988-231-2, Rainer Hampp Verlag, München und Mering 1997, 496 S., DM 69.80

Personalwirtschaftliche Problemstellungen werden häufig aus Sicht der Personalabteilung diskutiert. Im Vordergrund stehen dabei administrative sowie verfahrenstechnische Fragen. Die vorliegende Arbeit beschäftigt sich hingegen mit Fragen der Karrierepolitik als Teil der Unternehmenspolitik. Die Adressaten der Arbeit stellen neben den Arbeitnehmern daher nicht die Personalabteilung, sondern die Träger der Unternehmens- bzw. Personalpolitik dar - die Unternehmensleitung sowie das Kontrollorgan der Unternehmung. Diese haben ein zweistufiges Spiel im Sinne der nicht-kooperativen Spieltheorie zu lösen: Sie müssen mögliche Reaktionen der Arbeitnehmer auf verschiedene karrierepolitische Regelungen antizipieren, um daraufhin diejenigen Regelungen zu wählen, mit denen die Ziele der Karrierepolitik möglichst gut verfolgt werden. Angenommen wird hierbei, daß die Träger der Karrierepolitik ein Allokations- und Matchingziel sowie ein Anreizziel verfolgen.

Für die ökonomische Analyse der betrieblichen Karrierepolitik wird als Referenztheorien auf die neo-institutionalistische Theorie der Unternehmung und ausgewählte arbeitsökonomische Theorien zurückgegriffen. Zudem erfolgt ein Rückgriff auf die Fülle der existierenden empirischen Studien zur Personal- und Arbeitsökonomie. Die empirischen Ergebnisse werden insbesondere miteinbezogen, um die Rahmenbedingungen zu konkretisieren, die den karrierepolitischen Handlungsspielraum der Unternehmensleitung begrenzen.

no 46,80

3 8/13/12

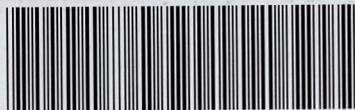
## Synthetische Unternehmensstichproben

Wie am Beispiel der Theorieentwicklung zum Arbeitsangebot gezeigt werden kann, setzt diese unter anderem die Verfügbarkeit von Daten zur Überprüfung dieser Theorien voraus. Zur Bearbeitung arbeitsnachfrageseitiger Fragestellungen sollte dabei idealerweise auf Unternehmens- oder Betriebsdaten zurückgegriffen werden, deren Verfügbarkeit und/ oder Verbreitung jedoch relativ gering ist. Von zentralem Interesse bei arbeitsnachfrageseitigen Fragestellungen dürfte der Einfluß von unterschiedlichem Nachfrageverhalten auf die Wertschöpfung im Unternehmen sein.

Mit synthetisch erzeugten Unternehmensdaten, die sich aus vorhandenen Aggregatdaten – wie beispielsweise der Arbeitsstättenzählung – konstruieren lassen, ist unter bestimmten Bedingungen ein Teil des Datendefizits zu beseitigen. Durch die so mögliche Verknüpfung verschiedener Aggregatdatensätze können auch Variable wie durchschnittliche Bruttowertschöpfung oder durchschnittliches Anlagevermögen mit in die empirische Analyse einbezogen werden. So weichen dann auch die Ergebnisse auf Basis der synthetischen Stichprobe (Unternehmensdaten) teilweise erheblich von den in der Literatur formulierten Ergebnissen auf Basis von Branchen- und/oder Individualdaten ab.

Synthetische Unternehmensstichproben, denen als Konstruktionsprinzip das Prinzip der maximalen Entropie zugrunde liegt, können aufgrund der ermutigenden und teilweise überraschenden Ergebnisse dieser Arbeit als ein möglicher Weg zu Beseitigung des Datendefizits in der empirischen Arbeitsmarktforschung betrachtet werden.

*Dr. Rainer Krütten*, Jg. 1962, Diplom-Kaufmann, 1996 Promotion zum Dr. rer. pol. an der Universität Trier, 1989-1993 wissenschaftlicher Mitarbeiter an der Universität Trier, Fachbereich Wirtschafts- und Sozialwissenschaften/ Mathematik, seit 1993 Beschäftigung in einer Wirtschaftsprüfungsgesellschaft.



206\$07894678

**Rainer Hampp Verlag**

ISBN 3-87988-255-X

München und Mering

ISSN 0948-6518

1997

SG: 17