

OLSEN, E.O. (1987); The Demand and Supply of Housing Service: A Critical Survey of the Empirical Literatur, in MILLS, E. S. (Eds.), Handbook of Regional and Urban Economics Elsevier Science Publishers B.V.

RÄTZER, E. (1987): „Mieterschutz und Wohnungsmarkt: Die Mietpreisbeschränkung im schweizerischen Missbrauchbeschluss“, Schweizerische Zeitschrift für Volkswirtschaft und Statistik, 123, 23-43.

ROSEN, K.T., SMITH, L.B. (1983): „Price-Adjustment Process for Rental Housing, the Natural Vacancy Rate“, American Economic Review, 73(4), 779-786.

6. QUELLEN

Statistisches Jahrbuch der Schweiz, verschiedene Ausgaben

Statistisches Jahrbuch des Kantons Zürich, verschiedene Ausgaben

Die Ermittlung und Prognose regionaler Beschäftigung mit dem ENTROP-Verfahren. Eine Anwendung auf Arbeitsmärkte in Ostdeutschland

Uwe Blien, Nürnberg und Friedrich Graef, Erlangen

Das ENTROP-Verfahren wurde zur Schätzung von Tabellen aus heterogenen Informationen entwickelt. Es eignet sich zur Bearbeitung vieler Fragestellungen der Regionalforschung aus den Bereichen Disaggregation von Daten, Stichprobengewichtung, Umschätzung von Regionalgliederungen, Erstellung von Regionalprognosen etc. Im vorliegenden Fall werden Testrechnungen zur Ermittlung regionaler Disparitäten auf Arbeitsmärkten Ostdeutschlands durchgeführt. Das Verfahren basiert auf dem Entropieoptimierungsprinzip, es ist in dem PC-Programm ENTROTAB implementiert.

Gliederung

1. Einführung
2. Aktualisierung von Verteilungen und Ermittlung regionaler Disparitäten
3. Einige Anmerkungen zur Anwendbarkeit der Entrop-Methode in der Regionalforschung
4. Der Entrop-Algorithmus

Literatur

Anhang

1. EINFÜHRUNG

Eine an die Regionalforschung des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) oft gerichtete Fragestellung ist die Deskription und Prognose der regionalen Beschäftigung. Letztere dienen vielfältigen analytischen Zwecken: Entwicklungsvergleiche für Regionen setzen differenzierte Informationen über die Bestände an Beschäftigung zu einem bestimmten Zeitpunkt voraus. Die Beurteilung eines bestimmten Entwicklungspfades wiederum ist von Interesse für die Gestaltung von Maßnahmen der Arbeitsmarkt- und der Strukturpolitik. Mittel für diese Maßnahmen sollen in den Brennpunkten regionaler Probleme ausgegeben werden. Dies erfordert eine exakte Ermittlung der regionalen Disparitäten.

Nun sind die dazu notwendigen Daten oft nicht mit einem hinreichenden Grad an Aktualität verfügbar. Manchmal hat man eine Totalerhebung für ein Jahr, während für die nachfolgenden Jahre nur stärker aggregierte Information vorliegt. Oder Regionaldaten sind nur mit bestimmter zeitlicher Verzögerung bekannt, ein Problem, das sich bei der Ermittlung von regionalen Disparitäten im IAB häufig stellt.

Im folgenden wird demonstriert, wie diese Schwierigkeiten durch Schätzungen mit dem sogenannten ENTROP-Verfahrens gelöst werden können. Diese Methode basiert auf dem Entropieoptimierungsprinzip und ist eine Verallgemeinerung des verbreiteten RAS-Verfahrens. Sie ist für die Schätzung von Matrizen (Tabellen) generell einsetzbar, wenn die Information, die über die zu ermittelnde Matrix vorliegt, durch lineare Gleichungen (z. B. Randsummen) und/oder Ungleichungen dargestellt werden kann und erlaubt die Berücksichtigung einer Vorgabetabelle, der die zu schätzende Matrix in einem noch zu präzisierenden Sinne so weit wie möglich „ähnlich“ ist. Damit eignet sich die Methode nicht nur zur Ermittlung regionaler Disparitäten, sondern auch generell zur

- Disaggregation von Daten
- Aktualisierung von Tabellen
- Ermittlung von Kontingenztabellen oder Übergangsmatrizen aus heterogenen Informationen (z. B. regionalen Input-Output-Tabellen)
- Gewichtung von Stichproben
- und zur Umschätzung von Daten einer Regionsgliederung auf eine andere.

Das ENTROP-Verfahren wurde für die Bildungsgesamtrechnung des IAB (Blien, Tessaring 1988, Fischer et al. 1993) entwickelt. Es erweist sich zusätzlich u. a. in der Regionalforschung als nützlich. Sein Anwendungsspektrum umfaßt auch den prognostischen Einsatz, wenn z. B. eine Tabelle, die eine Beschreibung regionaler Disparitäten enthält, auf der Basis geeigneter prognostischer Informationen in die Zukunft fortgeschrieben werden soll.

Im folgenden wird der Einsatz des ENTROP-Verfahrens für die Ermittlung und Prognose regionaler Disparitäten in einer Anwendung auf regionale Arbeitsmärkte in Ostdeutschlands dargestellt. Im vorliegenden Fall werden die Resultate von Testrechnungen vorgestellt, die eine Ermittlung der Verteilung von Beschäftigten auf Regionen und Branchen zum Gegenstand haben. Die Rechnungen lassen sich in der einfachen Variante als schlichte Aktualisierung von Verteilungen begreifen. In der komplexeren Variante können sie als Berechnung der Ausprägung regionaler Disparitäten auf der Grundlage heterogener Informationen verstanden werden. Schließlich sind sie als Variante einer Kurzfristprognose interpretierbar.

2. AKTUALISIERUNG VON VERTEILUNGEN UND ERMITTLUNG REGIONALER DISPARITÄTEN

Für die Analyse regionaler Arbeitsmärkte steht als besonders wertvolle Datenquelle die Beschäftigtenstatistik zur Verfügung. Diese weist allerdings Probleme auf (die im Falle des Prognosezwecks von Tassinopoulos 1996 und Bade 1994 diskutiert werden), insbesondere ist sie nur mit Verzögerung erhältlich, da sie auf den Meldungen der Betriebe und Krankenversicherungen beruht. Aus diesem Grunde sind Verfahren interessant, mit denen das Material unter Heranziehung externer Information aktualisiert werden kann. Wie funktioniert eine solche Methode und wie aussagefähig können die mit ihr erzielten Ergebnisse sein? Wie können die Schätzungen weiter verbessert werden, wenn heterogene Informationen verwendbar sind? Zu diesen Fragen werden im folgenden einige Antworten gegeben.

Die Untersuchung wird anhand von Daten der Beschäftigtenstatistik (Quartalsstatistik) über Ostdeutschland durchgeführt. Um die Vorgehensweise zu verdeutlichen, wird zunächst eine stark zusammengefaßte Gebietsgliederung verwendet, bei der nur 5 Regionen ausgewiesen werden. Diese sind im wesentlichen identisch mit den Bundesländern. Lediglich die Region „Nord“ besteht aus einer Zusammenfassung der Bundesländer Brandenburg und Mecklenburg-Vorpommern. Die Daten für Beschäftigte liegen in einer Gliederung nach 10 Wirtschaftszweigen vor.

Das Ziel der Rechnungen sei die Ermittlung einer Matrix, die eine Kombination von Branche und Region beinhaltet. Dazu sei zunächst angenommen, daß für beide Merkmale jeweils die Summen in den einzelnen Kategorien verfügbar sind. Diese Situation wird durch Tabelle 1 beschrieben, die die Randsummen für das Jahr 1994 enthält.

Weiterhin sei angenommen, daß die Kombination von Branche und Region für das Jahr vorher, d. h. für 1993, vorliegt. Tabelle 2 enthält diese komplette Verteilung. Dies ist eine einfache Situation. Zwar ist es nicht möglich, die komplette Tabelle durch die schlichte prozentuale Hochrechnung entweder der Zeilen- oder der Spaltensummen zu ermitteln, da eine Rechnung mit den Zeilensummen zu den bekannten Spaltensummen abweichende

Tabelle 1: Randsummen 1994

	Nord	Sachs.-A.	Sachsen	Thüringen	Berlin	Summe
Landw.						187199
Energie						38158
Verarb.						1043862
Bau						782352
Handel						580971
Verkehr						393836
Banken						102227
Dienst						1315930
Org./Hh.						185318
Körp.						750474
Summe	1537685	966940	1553625	990448	431629	5480327

Tabelle 2: Basistabelle 1993

	Nord	Sachs.-A.	Sachsen	Thüringen	Berlin	Summe
Landw.	80559	34807	44204	34583	2286	196439
Energie	56861	25248	52374	20019	7607	162109
Verarb.	247195	211994	351846	233880	59351	1104266
Bau	196837	127442	199191	127215	42682	693367
Handel	151407	99305	166112	98785	46085	561694
Verkehr	126512	78591	109081	63169	48797	426150
Banken	25631	16574	32571	18358	9529	102663
Dienst	333458	183818	365570	208205	140253	1231304
Org./Hh.	51048	23434	39884	26695	16674	157735
Körp.	243121	173949	190383	141004	83099	831556
Summe	1512629	975162	1551216	971913	456363	5467283

Ergebnisse erbringen würde (und umgekehrt). Für die Anpassung einer Tabelle an gegebene Randverteilungen steht jedoch das sogenannte RAS-Verfahren (vgl. Sen, Smith 1995: 373ff., Golan et al. 1996: 61, Wauschkuhn 1982) zur Verfügung. Die damit erzielten Ergebnisse sind mit Spalten- und Zeilensummen konsistent. Das hier verwendete ENTROP-Verfahren ist eine Verallgemeinerung von RAS. Werden nur Zeilen- und Spaltensummen vorgegeben, sind die beim ENTROP-Algorithmus ausgeführten Schritte mit denen des RAS-Verfahrens identisch (vgl. Blien, Graef 1991, 1995).

Der ausgeführte Algorithmus wird in Teil 4 näher erläutert. Sein Ergebnis, das in Form von Tabelle 3 vorliegt, hat die Eigenschaft, daß die vorgegebenen Randsummen exakt eingehalten werden. Gleichzeitig ist die geschätzte Tabelle der vorgegebenen in spezifischer Weise „ähnlich“, da in der Vorgabetabelle kleine Elemente auch in der resultierenden Tabelle klein sind. Tatsächlich hat der Algorithmus zwei weitere Eigenschaften, die für seine Praktikabilität unverzichtbar sind. Die Elemente der geschätzten Tabelle sind stets nichtnegativ (d. h. größer oder gleich Null). Außerdem bleiben Nullen erhalten, d. h. Elemente die in der Basistabelle Null sind, werden stets auch in der resultierenden Tabelle auf Null gesetzt.

Tabelle 3: Schätzung für 1994 auf der Basis der Randsummen

	Nord	Sachs.-A.	Sachsen	Thüringen	Berlin	Summe
Landw.	77376	32795	441719	33277	2031	187199
Energie	49042	21362	44387	17298	6069	38158
Verarb.	236713	199138	331068	224370	52573	1043862
Bau	224973	142884	223705	145664	45126	782352
Handel	158882	102223	171282	103850	44734	580971
Verkehr	118854	72427	100696	59453	42406	393836
Banken	25917	16440	32362	18596	8913	102227
Dienst	362259	195891	390239	226599	140943	1315930
Org./Hh.	60878	27414	46738	31894	18394	185318
Körp.	222790	156367	171429	129448	70440	750474
Summe	1537685	966940	1553625	990448	431629	5480327

Tabelle 4: Tatsächliche Werte 1994

	Nord	Sachs.-A.	Sachsen	Thüringen	Berlin	Summe
Landw.	75746	33525	41629	33904	2395	187199
Energie	46894	19202	46578	18851	6633	38158
Verarb.	240578	190627	335563	223417	53677	1043862
Bau	221265	153643	223554	140753	43137	782352
Handel	162949	103304	168009	103881	42828	580971
Verkehr	115416	72813	103498	59536	42573	393836
Banken	26048	16025	32661	18180	9313	102227
Dienst	360832	190947	391901	228397	143853	1315930
Org./Hh.	69123	23733	45387	32400	14675	185318
Körp.	218834	163121	164845	131129	72545	750474
Summe	1537685	966940	1553625	990448	431629	5480327

Die „Ähnlichkeit“ zwischen Basis- und Ergebnismatrix kommt dadurch zustande, daß beim RAS- und beim ENTROP-Verfahren ein Distanzmaß zwischen den beiden Tabellen minimiert wird, das relative Entropie genannt wird. Die Anwendung gerade dieses Distanzmaßes hat zur Konsequenz, daß die geschätzte Tabelle in einem spezifischen Sinne als die „wahrscheinlichste“ (Blien, Graef 1991) unter allen gelten kann, die die Randsummen von Tabelle 1 exakt einhalten.

Im vorliegenden Fall kann der Erfolg durch die Verwendung der bekannten Verteilung für das Jahr 1994 überprüft werden. Dazu braucht man ein Maß für den Unterschied zweier Tabellen. Die exakte Definition des hier verwendeten Maßes, das im folgenden als „relative Ähnlichkeit“ (sim) bezeichnet wird, kann Anhang 2 entnommen werden. Hier ist nur wichtig, daß der Wert dieses Maßes sim zwischen 0 und 1 liegt, wobei der Tabelle 2 (der Vorgabematrix 1993) der Wert 0 und der Tabelle 4 (der tatsächlichen Verteilung 1994) der Wert 1 (= 100 %) zugeordnet ist. Niedrige Prozentwerte zeigen demnach an, daß das Ergebnis unähnlich der tatsächlichen Verteilung ausfällt, hohe, daß es diese gut reproduziert. Es sei ausdrücklich darauf hingewiesen, daß das Maß keine Auskunft über die absolute Größe der Abweichungen gibt. Ein Wert von 90 % heißt keineswegs, daß die geschätzten Werte 90 % der tatsächlichen betragen. Im vorliegenden Kontext sind die absoluten Abweichungen viel kleiner.

Entsprechend dem definierten Kriterium ist das Ergebnis ziemlich gut ausgefallen, da der Wert für die relative Ähnlichkeit 87 % beträgt. Danach reproduziert die geschätzte Tabelle weit eher die tatsächliche Verteilung als die Vorgabematrix. Das gute Ergebnis wurde erzielt, obwohl sich Ostdeutschland in den Jahren 1993 und 1994 in einer Umbruchsituation befindet, man also nicht von Stabilität der regionalen Disparitäten ausgehen kann.

Wie erwähnt kann diese Schätzung bereits mit RAS ausgeführt werden. Gegenüber RAS besteht der Vorzug von ENTROP in der größeren Flexibilität. Oft hat man zusätzliche Informationen über die zu schätzende Matrix, die man gerne im Schätzprozeß berücksichtigen würde. Dies ist bei RAS meist nicht möglich, lediglich in verschiedenen Grenzfällen kann mit „Tricks“ gearbeitet werden.

Im Falle von ENTROP können beliebige zusätzliche Informationen in die Schätzung eingebaut werden, wenn sie in die Form linearer Gleichungen oder Ungleichungen trans-

formiert werden können. Die Aufnahme von Ungleichungen ermöglicht z. B. die Berücksichtigung qualitativer Informationen, die nur mit einer bestimmten Bandbreite quantifizierbar sind.

Tabelle 5: Zusatzinformationen über 1994

	Nord	Sachs.-A.	Sachsen	Thüringen	Berlin	Summe
Landw.					2395	187199
Energie		19202	> 72000			38158
Verarb.		1267500 ± 500				1043862
Bau						782352
Handel						580971
Verkehr	< 115450					393836
Banken						102227
Dienst	< 552000		Verhältnis > 1,7 zu 1			1315930
Org./Hh.	> 69000				< 14700	185318
Körp.						750474
Summe	1537685	966940	1553625	990448	431629	5480327

Tabelle 6: Schätzung unter Berücksichtigung der Zusatzinformationen.

	Nord	Sachs.-A.	Sachsen	Thüringen	Berlin	Summe
Landw.	77465	32874	41403	33062	2395	187199
Energie	46956	19202	47122	18384	6494	38158
Verarb.	236686	200564	330114	223978	52520	1043862
Bau	224932	143898	223046	145400	45077	782352
Handel	159484	102740	170432	103452	44864	580971
Verkehr	115450	73759	101525	60010	43092	393836
Banken	26021	16527	32209	18530	8941	102227
Dienst	358109	193891	392696	228285	142949	1315930
Org./Hh.	69094	26432	44614	30479	14700	185318
Körp.	223488	157054	170466	128867	70598	750474
Summe	1537685	966940	1553625	990448	431629	5480327

Im vorliegenden Fall wurde angenommen, daß über einzelne Matrixelemente von Tabelle 4 zusätzliche Informationen erhältlich sind. Die vorhandenen Informationen wurden in Tabelle 5 eingetragen. Der Wert von 19202 in dem in der zweiten Spalte und zweiten Zeile stehenden Tabellenzelle bedeutet, daß die Anzahl der Beschäftigten im Energiebereich Sachsen-Anhalts exakt diesen Wert aufweisen soll. Der von der zweiten bis zur vierten Spalte und von der dritten bis zur vierten Zeile reichende „Kasten“, mit dem Wert 1267500 ± 500 soll bedeuten, daß in den Branchen „Verarbeitendes Gewerbe“ und „Energie“ in den Länder Sachsen-Anhalt, Sachsen und Thüringen (im Jahre 1994) 1267500 Menschen arbeiten, wobei eine Unsicherheit von plus oder minus 500 besteht.

Tabelle 6 zeigt das mit ENTROP erzielte Ergebnis. Erneut weist es exakt die Randsummen von Tabelle 1 auf. Darüberhinaus sind alle in Tabelle 5 eingetragenen Vorgaben berücksichtigt, wie man sich leicht überzeugen kann, und die Tabelle 6 hat eine zu Tabelle 2 oder Tabelle 4 ähnliche Struktur. Der Erfolg der Schätzung kann mit dem Kriterium „relative Ähnlichkeit“ *sim* überprüft werden. Das Abstandsmaß zeigt eine Verbesserung der Schätzung, sein numerischer Wert beträgt 94 %. Damit liegt das Schätzergebnis sehr nahe an der tatsächlichen Verteilung.

Es ist nicht überraschend, daß die Integration von Informationen über die „wirkliche“ Verteilung von Beschäftigten das Schätzergebnis weiter gegenüber der RAS-Schätzung verbessert. Allerdings ist wichtig, daß ENTROP überhaupt erst das Instrument zur Verfügung stellt, mit dem eine flexible Berücksichtigung verschiedener Informationsquellen möglich ist.

Dies kann in gleicher Weise bei größeren Tabellen demonstriert werden, die in Anhang 1 dokumentiert sind. Diese Matrizen enthalten die bereits bekannte Kategorisierung von Branchen, nur der Energiebereich wurde mit dem verarbeitenden Gewerbe zusammengefaßt. Die verwendete Regionsgliederung ist jedoch wesentlich feiner, da Arbeitsamtsbezirke (in unbereinigter Form) zugrundegelegt werden. Tabelle A1 entspricht der Tabelle 1, sie ist die Basismatrix der Schätzung und zeigt die Verteilung der Beschäftigten im Jahre 1993. Tabelle A2 entspricht Tabelle 4, sie gibt die wirklichen Werte an für 1994 an. Tabelle A3 ist eine Matrix mit Randsummen und Zusatzinformationen für das Jahr 1994. Allerdings wurde in diesem Fall von stärker aggregierter regionaler Gliederung ausgegangen, d. h. angenommen, daß für 1994 keine detailliertere Regionalinformation vorhanden ist.

Tabelle A4 zeigt das mit den Informationen aus den Tabellen A1 und A3 erzielte Ergebnis. Es wird deutlich, daß eine Schätzung mit ENTROP in diesem Fall eine *Disaggregation der zusammengefaßten Regionalinformation* bedeutet. Von einer Regionalgliederung wird auf eine andere, eine feinere, umgeschätzt. Weitere Ergebnistabellen sind A4, bei der nur auf Basis von disaggregierten Randsummen geschätzt wurde und A5, bei der die Zusatzinformation von A2 zusätzlich eingebaut wurde.

Bei den einzelnen Schätzungen sind die Werte der relativen Ähnlichkeit *sim* angegeben. In diesem Falle treten etwas kleinere Werte auf, da die einzelnen Tabellenelemente in den größeren Matrizen durch die Randsummenvorgaben nicht mehr derart eng vorgegeben sind, wie bei den kleineren Tabellen. Erneut zeigt sich jedoch, daß die Berücksichtigung zusätzlicher Vorgaben das mit ENTROP ermittelte Ergebnis verbessert.

Die Berechnung der Schätzungen erfolgte mit einem für PCs unter MSDOS/PCDOS erstellten Programm ENTROTAB, das auch eine eigene Tabellenverwaltung und eine übersichtliche Eingabe der Randsummen und Zusatzinformationen bietet. Eine Windows-Version dazu ist in Vorbereitung.

3. EINIGE ANMERKUNGEN ZUR ANWENDBARKEIT DER ENTROP-METHODE IN DER REGIONALFORSCHUNG

In der Regionalforschung bieten sich vielfältige Anwendungsmöglichkeiten für das ENTROP-Verfahren, da man hier oft das Problem hat, daß die vorhandene Information nicht oder nur unvollkommen regionalisiert ist. In diesen Fällen kann die Methode unter bestimmten Bedingungen zur Disaggregation eingesetzt werden. Dabei muß der Anwender natürlich stets darauf achten, daß eine derartige Vorgehensweise sinnvoll ist. Eine wesentliche Bedingung für den Einsatz ist Strukturkonstanz. Nur dann kann die Struktur einer Basismatrix an neue Randverteilungen und sonstige Informationen angepaßt werden. Wir haben allerdings bereits gesehen, daß diese Bedingung großzügig ausgelegt werden kann. Im Fall der neuen Länder, die sicherlich von Strukturbrüchen betroffen sind, ergab sich dennoch ein „realistisches“ Ergebnis. Ansonsten ist der Einsatz der ENTROP-Methode immer dann sinnvoll, wenn die Randsummen und sonstigen Gleichungs- und Unglei-

chungsvorgaben bereits derart informativ sind, daß sie das Ergebnis schon weitgehend determinieren.

Unter Beachtung dieser Bedingungen ergibt sich ein weites Spektrum von Anwendungen im Bereich der Regionalforschung. Zu denken ist z. B. an die Schätzung von regionalen Input-Output-Matrizen oder von Mobilitätsmatrizen.

Außerhalb der Regionalforschung wurde das ENTROP-Verfahren bereits bei verschiedenen Problemstellungen im sozial- und wirtschaftswissenschaftlichen Bereich im IAB erfolgreich angewendet. Z. B. wurden bei bildungsstatistischen Analysen mehrdimensionale Tabellen geschätzt, die mehrere zehntausend Felder aufweisen. Für die Schätzungen wurden jeweils mehrere tausend Vorgaben spezifiziert. Die Flexibilität des Verfahrens wird auch daraus deutlich, daß eine Vorgabetabelle nicht vorhanden sein muß, aber vorhanden sein kann.

In der Regionalforschung des IAB werden derzeit verschiedene Anwendungen vorbereitet, die durch die hier dargestellten Tests schon nahegelegt werden. Das sogenannte „Jahreszeitraummaterial“ (vgl. z. B. Blien, Koller, Schiebel 1991, Koller 1990), das alle Meldungen enthält (ebenso die sog. „Historikdatei“, vgl. Cramer, Maier 1991), kann jeweils erst verwendet werden, wenn die Meldungen eingetroffen sind, d. h. frühestens mit einer Verzögerung von zwei Jahren. Schneller (ab ca. 9 Monaten) liegt die Quartalsstatistik der Beschäftigtenstatistik vor. Bei dieser findet eine bestimmte Auffüllung fehlender Information statt, die durch das sogenannte „Abschneiderverfahren“ beschrieben wird (Cramer 1985, Cramer, Maier 1991). Damit ist ein bestimmter, normalerweise kleiner Fehler verbunden, der sich i. d. R. lediglich bei kleinräumlicher Gliederung von Gebietseinheiten störend auswirken kann. Aus diesem Grunde könnte der Versuch gemacht werden, das Jahreszeitraummaterial durch die Verwendung der Quartalsstatistik zu aktualisieren. Dies wäre in der Weise möglich, die oben beispielhaft demonstriert wurde. Hier wird die Basistabelle aus dem Jahreszeitraummaterial entnommen und die aktuellen Randverteilungen aus der Quartalsstatistik.

Eine andere Möglichkeit liegt im Prognoseeinsatz, zu dessen Beschreibung hier etwas weiter ausgeholt sei. Prognosen sind vor allem für praktische Verwendungen in der Arbeitsmarktpolitik von Interesse. Politische Akteure haben einen spezifischen Bedarf an prognostische Ausblicken in eine wahrscheinliche Zukunft. Sie wollen knappe Budget-

mittel insbesondere dorthin steuern, wo Einbrüche zu erwarten sind, damit sie schon in der Gegenwart gegensteuern können, um eine prognostizierte Krise erst gar nicht eintreten zu lassen. Aus diesen Gründen wird bei den regelmäßig stattfindenden Neuabgrenzungen der Fördergebiete der Gemeinschaftsaufgabe "Verbesserung der regionalen Wirtschaftsstruktur" (vgl. Blien, Koller, Schiebel 1991, Hirschenauer 1994, Zarth 1991) auch ein Indikator herangezogen, der auf einer regionalen Beschäftigungsprognose beruht (vgl. Bade 1991, 1994).

Nun sind Prognosen eine schwierige Angelegenheit. Abgesehen davon, daß jede Zukunft naturgemäß unsicher sein muß, stellen sich bei Prognosen ökonomischen Inhalts zwei prinzipielle Schranken: Die eine der beiden besteht in der grundsätzlichen Entscheidungsfreiheit des Menschen, die eine vollständig deterministische Sozialwissenschaft ausschließt. Eine wesentliche Erfahrung jüngerer Forschung zu der Stabilität von Präferenzen, die von Ökonomen, Soziologen und Psychologen betrieben wird, erbrachte, daß Präferenzen oft zeitinstabil sind (vgl. Conlisk 1996 für einen Überblick). Subjekte entscheiden sich oft immer wieder neu und zum Teil erst in der Situation, in der sie handeln.

Die andere der beiden Schwierigkeiten besteht darin, daß eine Marktwirtschaft geradezu auf den dezentralen Prozessen der Entscheidungsfindung der Wirtschaftssubjekte beruht. Es erweist sich immer erst nachträglich, ob ihre Pläne wirklich kompatibel sind. Schon aus diesen beiden Gründen können Prognosen nur von begrenzter Zuverlässigkeit sein. Insbesondere Wendepunkte der Entwicklung widerstehen auch besonders komplizierten und raffinierten Verfahren (Campbell 1993, Campbell, Ormerod 1992a, 1992b).

Andererseits gibt es Gründe dafür, daß im Falle von Prognosen mit kurzfristigem Zeithorizont, und nur solche sollen hier diskutiert werden, die Beschäftigung mit einiger Genauigkeit vorhergeschätzt werden kann. Der Arbeitsmarkt reagiert auf Veränderungen der Produktion und diese auf den Auftragseingang und auf Erwartungen der Wirtschaftssubjekte. In dieser Reaktionskette treten Verzögerungen auf, die regelmäßig für prognostische Zwecke genutzt werden, z. B. in der kurzfristigen, globalen Arbeitsmarktprognose des IAB (vgl. Autorengemeinschaft 1988).

Unter bestimmten Umständen erscheint es möglich, in einer Top-Down-Vorgehensweise bei der Prognose einzelne Aggregatwerte mit größerer Genauigkeit vorauszuschätzen als detaillierte regionale Aufgliederungen. Damit ist erneut ein Analysedesign

vorgezeichnet, daß sich stark an den hier vorgestellten Beispielen orientiert. Es ist möglich, die Entwicklung von einzelnen Branchen und Bundesländern zu prognostizieren. Dies kann z. B. auf der Grundlage der Auftragseingänge oder anderer „leading indicators“ erfolgen. Die Ergebnisse dieses ersten Prognoseschritts können dann auf mit dem ENTROP-Verfahren auf kleine regionale Einheiten übertragen werden. Die Verteilung der Beschäftigung in einer Basisperiode dient als Vorgabematrix. Zusätzlich können in die ENTROP-Schätzung vielfältige qualitative Urteile über regionale Sonderentwicklungen in Form von (Un-)gleichungsrestriktionen eingebaut werden. Ebenso sind anstelle fixer Randsummenvorgaben auch Bandbreiten über Verhältnisse zwischen Branchen oder Regionen möglich. So kann man vorgeben, daß der Prognosewert für Sachsen zwischen dem 1,600- und dem 1,625-fachen des Wertes von Sachsen-Anhalt liegen soll und anderes mehr.

4. DER ENTROP-ALGORITHMUS

Das bei der Realisierung des ENTROP-Verfahrens angewendete Entropieoptimierungsprinzip, das der Methode den Namen gegeben hat, ist in den Natur- und Ingenieurwissenschaften weit verbreitet. Es wird z. B. auch in der Medizin, nämlich bei der Computertomographie, angewendet. Dabei wird aus Röntgenbildern, die aus verschiedenen Winkeln aufgenommen werden, das Schnittbild eines Organs (in weiteren Schritten bei fortentwickelten Methoden auch eine dreidimensionale Darstellung) zusammengesetzt. Die Analogie zur Schätzung einer Tabelle besteht darin, daß man sich ein gerastertes Bild als eine Tabelle mit sehr vielen Zeilen und Spalten vorstellen kann, in deren Zellen die Grauwerte für einzelne Bildpunkte enthalten sind. Den von verschiedenen Seiten aufgenommenen Röntgenbildern entsprechen die Zeilen- und Spaltensummen der Tabelle. Die Leistungsfähigkeit der Entropieoptimierung wird dadurch belegt, daß mit ihr tatsächlich ein der Wirklichkeit entsprechendes Schnittbild des betreffenden Organs, einschließlich eventuell vorhandener krankhafter Veränderungen, produziert werden kann.

In der Regionalwissenschaft ist das Entropieoptimierungsprinzip seit Wilsons Arbeiten (1970), die ihrerseits auf Shannon (vgl. Shannon, Weaver 1949) zurückgehen, vor allem im Zusammenhang von Gravitationsmodellen (vgl. für einen modernen Überblick Sen,

Smith 1995) bekannt. Hier sei jedoch betont, daß die inhaltlichen Implikationen dieser Modelle im vorliegenden Kontext nicht vorausgesetzt werden. Bei der Entropieoptimierung handelt es sich um ein allgemeines Schätzprinzip, das für die vorliegende Fragestellung genutzt werden kann. Im folgenden werden das Optimierungsproblem, auf das hin der ENTROP-Algorithmus entwickelt wurde, formal dargestellt und die einzelnen Schritte der Methode ebenfalls exakt angegeben. Leser, die nicht an mathematischen Details interessiert sind, können diese Ausführungen überspringen. Hier sei noch angemerkt, daß die verwendete Distanzfunktion, die relative Entropie, von Kullback (1959) eingeführt wurde. Sie wird in verschiedenen Varianten verwendet und firmiert unter verschiedenen Namen (z. B. als "cross-entropy", vgl. Golan, Judge, Miller 1996).

ENTROP kann auf vieldimensionale Tabellen angewendet werden. Zur Vereinfachung der Formeln stellen wir Mehrfachindizes für die einzelnen Tabellendimensionen durch einen einzigen Buchstaben (i, k, \dots) dar und bezeichnen mit M die Menge der auftretenden Mehrfachindizes. Eine Tabelle ist dann eine durch die Indizes k aus M indizierte Gruppe von Zellenbelegungen: $x = (x_k; k \in M)$

Die Tabellenschätzung kann als Optimierungsproblem in dem Sinne begriffen werden, daß eine Abstandsfunktion

$$F(x; u) = \sum_{k \in M} u_k f\left(\frac{x_k}{u_k}\right)$$

zwischen einer Basismatrix u und einer zu schätzenden Matrix x unter R Nebenbedingungen der Form

$$b'_r \leq g_r(x) \leq b''_r$$

für alle $r = 1, 2, \dots, R$ und der Nichtnegativitätsforderung $x_k \geq 0$ für alle $k \in M$ minimiert wird. Die Funktionen g_r sind dabei lineare Funktionen der Zellenwerte x_k :

$$g_r(x) = \sum_{k \in M} a_{rk} x_k$$

Die meisten üblicherweise verfügbaren Informationen über die zu schätzende Tabelle lassen sich mittels derartiger Informationen darstellen.

In die Abstandsfunktion F geht ein spezielles Teilmaß f ein, das noch näher definiert werden muß. Diese Funktion f soll bewirken, daß im Minimum die Quotienten x_k/u_k im

Idealfall gleich 1 sind, bzw. so nahe bei 1 liegen, wie es die Nebenbedingungen nur zulassen. Sie sollte daher die Form der in Abb. 1 skizzierten Funktionen besitzen, d. h. eine konvexe Funktion mit Minimalstelle $t = 1$ sein. Die Gewichtungsfaktoren u_k bewirken, daß bei gleicher prozentualer Abweichung der Zellenwerte der Schätztable von denen der Basistabelle die Zellen mit der größeren Belegungszahl bei der Abstandsmessung stärker ins Gewicht fallen.

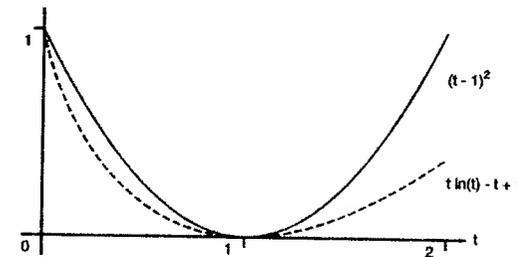


Abb. 1: Die Funktionen f_q und f_e

Es bieten sich im wesentlichen zwei Funktionen an. Dabei handelt es sich einmal um die quadratische Abweichung $f_q(t) = (t - 1)^2$ und die Funktion $f_e(t) = t \ln(t) - t + 1$, die in F eingesetzt die *relative Entropie* ergibt.

Welche dieser Funktionen benutzt wird, ist weniger eine Frage des gewünschten Resultats, weil sich die Minimalstellen in der Praxis kaum unterscheiden. Ein formaler Beweis dafür findet sich z. B. bei Kadas und Klafsky (1976), wo gezeigt wird, daß im Falle einer Rechnung mit durch die Tabellensummen normierten Werten, eine Minimierung beider Funktionen zu approximativ identischen Ergebnissen führt. Ansonsten sei für die Diskussion verschiedener Verfahren zur Tabellenschätzung auf Wauschkuhn (1982) verwiesen.

Für die Verwendung der quadratischen Abweichung spricht die Verwandtschaft mit der χ^2 -Statistik. Der theoretische Hintergrund der Verwendung der relativen Entropie besteht u. a. im Wilsonschen Entropiekonzept. Für die Verwendung der relativen Entropie spricht, daß sich zeigen läßt, daß ihre Optimierung der Maximierung der „Wahrscheinlichkeit“ der Ergebnistabelle entspricht, wenn sich letztere durch eine Multinomialverteilung charakte-

risieren läßt (vgl. die Darstellung des Beweises bei Blien, Graef 1995). Bei der Entropiefunktion spielen zudem Ausreißer keine derart problematische Rolle wie bei der quadratischen Funktion.

Aus praktischer Sicht hat die Entropiefunktion schließlich den Vorteil, daß man sich beim Berechnungsverfahren nicht darum kümmern muß, daß die iterativ berechneten Tabellenwerte nichtnegativ bleiben, da dies bei dieser Funktion automatisch gesichert ist. Der ENTROP-Algorithmus benutzt diese Funktion. Er wurde ursprünglich als eine Verallgemeinerung eines aus der Computertomographie bekannten Verfahrens entwickelt. Die wesentliche Ergänzung war die zusätzliche Berücksichtigung einer Basistabelle. Im Nachhinein stellte sich heraus, daß es sich beim ENTROP-Verfahren um einen Spezialfall einer Klasse von Optimierungsproblemen handelt, die erstmals 1967 von dem russischen Mathematiker Bregman studiert worden war.

Nach dem Satz von Kuhn, Tucker und Karoush ist die Struktur des Optimierungsproblems bekannt. Es ist

$$x_k = u_k e^{-\sum_r \lambda_r a_{rk}}$$

für alle $k \in M$, wobei die Multiplikatoren λ_r eine durch die jeweilige Nebenbedingung bestimmte Vorzeichenbedingung erfüllen müssen:

$$\lambda_r = 0, \text{ falls } b_r' < g_r(x) < b_r''$$

$$\lambda_r \geq 0, \text{ falls } g_r(x) = b_r'$$

$$\lambda_r \leq 0, \text{ falls } g_r(x) = b_r''$$

Das Schema des ENTROP-Algorithmus gleicht jenem des RAS-Verfahrens (das übrigens mit dem IPF-Algorithmus aus den loglinearen Modellen identisch ist). Es wird iterativ eine Folge von Tabellen $x = (x_k)$ mit Komponenten der obigen Form berechnet, die gegen die Lösung des Optimierungsproblems konvergiert.

Als Startwerte werden alle λ_r gleich Null gesetzt bzw. die $x_k = u_k$ für alle k . In jedem Iterationsschritt werden sodann alle Nebenbedingungen in einer festgelegten Reihenfolge durchlaufen, wobei jeweils durch Änderung des Multiplikators λ_r die Funktion $g_r(x)$ unter Beachtung der Vorzeichenbedingung an die Schranken b_r' bzw. b_r'' „herangeschoben“ wird. Diese Art der Projektion heißt nach ihrem Erfinder Bregmansche D-Projektion bzw.

hier Entropieprojektion. Aus der Sicht der Optimierungstheorie handelt es sich bei dieser Vorgehensweise um ein Koordinatenanstiegsverfahren zur Lösung des dualen Optimierungsproblems. Mit der Variation $\lambda_r - \delta$ ergibt sich die Tabelle $x(\delta)$ mit den Komponenten

$$x_k(\delta) = u_k e^{-\sum_j \lambda_j a_{jk} + \delta a_{rk}} = x_k e^{\delta a_{rk}}$$

Als Funktion der Variable δ ist $g_r(x(\delta)) = \sum_{k \in M} a_{rk} x_k e^{\delta a_{rk}}$ monoton steigend. Ist $g_r(x(0)) < b_r'$, so ist $\lambda_r \geq 0$ und man läßt δ so lange wachsen, bis entweder $g_r(x(\delta)) = b_r'$ oder $\lambda_r - \delta = 0$. Entsprechend verfährt man im Fall $g_r(x(0)) > b_r''$. Dann wird $x = x(\delta)$ gesetzt und die nächste Nebenbedingung bearbeitet. Falls überhaupt eine Lösung des Optimierungsproblems existiert, konvergiert die nach dem obigen Schema berechnete Folge von Tabellen gegen diese (eindeutig bestimmte) Lösung.

Das ENTROP-Verfahren ist nicht das schnellste mögliche. Da es aber die funktionale Form der relativen Entropie adäquat nutzt, ist es i. d. R. viel schneller als die üblichen allgemein anwendbaren numerischen Verfahren der nichtlinearen Optimierung (z. B. stochastische Optimierung). Es ist hinreichend schnell, um selbst große Probleme wie in der Bildungsgesamtrechnung des IAB (vierdimensionale Matrizen mit ca. 30000 Elementen, ca. 2000 Restriktionsvorgaben) leicht berechnen zu können. Es hat außerdem zwei Vorteile:

- Der Hauptspeicherbedarf ist relativ klein. Da bei jedem Rechenschritt nur die Nicht-Null-Komponenten a_{rk} der r -ten Nebenbedingung und nur die zugehörigen Tabellenwerte x_k benutzt werden, lassen sich auch große Tabellen bei relativ geringem Speicherausbau bearbeiten, natürlich auf Kosten der Rechenzeit.

- Die Methode bietet zusätzliche Möglichkeiten der Ablaufkontrolle. Anhand der Multiplikatoren λ_r und der Abweichung der Funktionswerte $g_r(x)$ von den Schranken läßt sich auf einfache Weise feststellen, ob und welche Nebenbedingungen sich gegenseitig ausschließen. Derartige Probleme treten häufig auf, wenn viele Informationen aus unterschiedlichen Quellen in die Nebenbedingungen einfließen.

Literaturverzeichnis:

- Autorengemeinschaft (1988): "Kurzfristige Arbeitsmarktanalysen und -prognosen. Konzepte, Methoden, Ergebnisse und Probleme", in: Mertens (1988): 226-245
- Bade, Franz-Josef (1991): "Regionale Beschäftigungsprognose 1995", in: Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 24/1, S. 25 - 44
- Bade, Franz-Josef (1994): "Die Beschäftigungsentwicklung bis zum Jahr 2000 in den Regionen der Bundesrepublik Deutschland. Prognose 1992-2000 und ex post-Kontrolle der Prognose 1987-1992", in: Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 27/2: 137-151
- Blien, Uwe; Graef, Friedrich (1991): "Entropieoptimierungsverfahren in der empirischen Wirtschaftsforschung. Die Ermittlung von Tabellen aus unvollständigen Informationen, dargestellt am Beispiel der Input-Output-Analyse", in: Jahrbücher für Nationalökonomie und Statistik 208/4: 399-413
- Blien, Uwe; Graef, Friedrich (1995): "Entropy Optimizing Methods for the Estimation of Matrices", paper presented at the Whitsun Meeting of the German Statistical Association, Dresden
- Blien, Uwe; Hirschenauer, Franziska (1995): "Weder Konvergenz noch Mezzogiorno - Zur Entwicklung regionaler Disparitäten in Ostdeutschland", in: Beschäftigungsobservatorium Ostdeutschland
- Blien, Uwe; Koller, Martin; Schiebel, Winfried (1991): "Indikatoren zur Neuabgrenzung der Förderregionen", in: Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 24/1: 1-24
- Blien, Uwe; Tessaring, Manfred (1988): "Die Bildungsgesamtrechnung des IAB. Ein Kohortenkonzept zur Analyse von Bildung und Erwerbstätigkeit", in: Mertens (1988), S. 144ff.
- Bregman, L. (1967): "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming", in: Urban Studies/Soziologische Revue Comp. Math. and Math. Phys. Vol. 7/3, S. 200ff.
- Campbell, Michael (1993): "On forecasting unemployment", unpublished paper, London
- Campbell, Michael; Ormerod, Paul (1992): "A Connectionist Autoregressive Model of the UK Economy", unpublished paper presented at the PASE 1992 in Prag

- Campbell, Michael; Ormerod, Paul (1992): "A Connectionist Vector Autoregressive Model of the UK Economy", in: Neural Network World 2/6: 571-582
- Conlisk, John (1996): "Why Bounded Rationality?", in: Journal of Economic Literature 34/2: 669-700
- Cramer, Ulrich (1985): "Genauigkeitsprobleme der Beschäftigtenstatistik", in: Allgemeines Statistisches Archiv 1
- Cramer, Ulrich; Majer, Wolfgang (1991): "Ist die Beschäftigtenstatistik revisionsbedürftig?", in: Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 24/1: 81-90
- Fischer, G.; Hensel, R.; Reinberg, A.; Reuschel, H.; Tessaring, M. (1993): "Bestand und Bewegung im Bildungs- und Beschäftigungssystem der Bundesrepublik Deutschland - Ergebnisse der Bildungsgesamtrechnung für die alten und neuen Bundesländer" (Beiträge zur Arbeitsmarkt- und Berufsforschung 170), Nürnberg: Bundesanstalt für Arbeit
- Golan, Amos; Judge, George; Miller, Douglas (1996): "Maximum Entropy Econometrics. Robust Estimation with Limited Data", New York etc.: John Wiley and Sons
- Hirschenauer, Franziska (1994): "Indikatoren zur Neuabgrenzung des regionalpolitischen Fördergebiets 1993", in: Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 27/2: 108-129
- Kadas, S. A.; Klafsky, E. (1976): "Estimation of the Parameters in the Gravity Model for Trip Distribution: a New Model and Solution Algorithm", in: Regional Science and Urban Economics Vol. 6, S. 439ff.
- Koller, Martin (1990): "Zur Erfolgskontrolle der regionalen Struktur- und Arbeitsmarktpolitik: Vorschläge zur Auswahl und Beobachtung strukturschwacher Regionen", in: Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 23/2, S. 227-254
- Koller, Martin et al. (1994): "REGARIS. Regionales Arbeitsmarktinformationssystem des IAB" (Projektbeschreibung des IAB), Nürnberg
- Kullback, Salomon (1968): "Information Theory and Statistics", zweite Auflage, New York: Dover
- Mertens, Dieter (1988) (Hrsg.): "Konzepte der Arbeitsmarkt- und Berufsforschung. Eine Forschungsinventur des IAB", 3. erweiterte und überarbeitete Auflage (Beiträge zur Arbeitsmarkt- und Berufsforschung 70), Nürnberg

Sen, Ashish; Smith, Tony (1995): "Gravity Models of Spatial Interaction Behavior", Berlin etc.: Springer

Shannon, Claude E.; Weaver, Warren (1949): "The Mathematical Theory of Communication", Urbana: University of Illinois Press

Tassinopoulos, Alexandros (1996): "Regionale Beschäftigungsprognose 1996/97 für die alten Bundesländer", Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 3

Wauschkuhn, Udo (1982): "Anpassung von Stichproben und n-dimensionalen Tabellen an Randbedingungen. Bewertung und Vergleich existierender Verfahren" (Berichte der Gesellschaft für Mathematik und Datenverarbeitung Nr. 138), München, Wien: R. Oldenbourg

Wilson, Alan G. (1970): "Entropy in urban and regional modelling", London: Pion Ltd.

Zarth, Michael (1991): "Neuordnung der Regionalförderung im Rahmen der Gemeinschaftsaufgabe "Verbesserung der regionalen Wirtschaftsstruktur"", in: Informationen zur Raumentwicklung 9-10: 539-556

Anhang I: Schätzung der Verteilung von Beschäftigten nach Arbeitsamtsbezirken

Tabelle A1: Basistabelle 1993

Bezirk	Land	En/Gew	Bau	Handel	Verkehr	Banken	Dienst	Org./Hh	KöRp.	Summe
Neubdb	11610	22559	21091	16915	10703	3046	33120	5694	25832	150570
Rostock	8415	31880	21062	20798	20202	3154	50047	8331	21808	185697
Schwer	10894	32951	23424	18320	14645	3620	38395	7478	26991	176718
Strals	5428	18514	14820	11671	10256	1586	33006	4042	16196	115519
Cottbus	9457	74456	28214	21400	19795	3543	43541	4174	34870	239450
Ebersw	6735	21629	15907	10400	7649	1541	22013	3373	19996	109243
Fkfrt/O	7162	27749	18303	13077	11354	2225	29800	6244	24849	140763
Neurup	11790	33760	20607	15289	9799	2194	28984	4139	24178	150740
Potsdam	9068	40558	33409	23537	22109	4722	54552	7573	48401	243929
Dessau	2746	23442	11288	9408	6671	1459	15689	3167	12821	86691
Halbers	4188	23492	11866	10329	6692	1449	20039	1360	15950	95365
Halle	2387	40719	24437	19077	22660	2856	48733	3420	35847	200136
Magdeb	6907	60668	32437	27737	21665	5298	46715	8197	50281	259905
Mersebg	5023	44629	17729	12569	7306	1639	19099	1843	20061	129898
Sangerh	2906	18355	9799	8193	4228	1564	12554	2754	15635	75988
Stendal	7407	12378	12519	7898	7101	1512	12208	1948	16956	79927
Wittenb	3243	13559	7367	4094	2268	797	8781	745	6398	47252
Altenbg	1834	8291	5307	4015	1724	557	5883	481	4888	32980
Annabg	3563	32773	12193	9156	4603	1737	22876	1530	14589	103020
Bautzen	7273	60422	31003	19327	13107	3773	41609	4632	30226	211372
Chemn	4336	54806	25351	30271	14952	5441	45627	7231	31011	219026
Dresden	4127	54777	26869	30281	24645	6837	75656	11253	37316	271761
Leipzig	6978	90622	47441	42034	31548	8391	103532	8266	34186	372998
Oschatz	3700	14910	7927	6249	2371	878	10113	963	7643	54754
Pirna	6843	34051	18309	9241	5130	2332	26544	1584	11695	115729
Plauen	2503	30981	13401	8327	5846	1432	19509	2447	10475	94921
Riesa	3047	22587	11390	7211	5155	1193	14221	1497	8354	74655
Zwickau	3491	39683	18955	15540	9535	2379	30384	3100	16405	139472
Erfurt	6837	41780	26321	23030	17847	5374	47497	6298	36047	211031
Gera	4090	27918	17341	11871	7258	2321	22258	4214	16741	114012
Gotha	5012	32020	15307	13588	7241	1841	24657	5339	14305	119310
Jena	4621	41230	14867	11341	6917	2167	31624	2729	15577	131073
Nordhn	4027	22330	13670	9351	4848	1370	20418	1810	16147	93971
Suhl	6505	48938	20754	14064	9523	2906	31367	3205	25782	163044
Berlin 1	280	21034	10150	24081	34610	8612	71122	10685	38433	219007
Berlin 2	1045	15445	10923	8464	3393	357	31650	3017	18948	93242
Berlin 3	524	13283	11217	8370	2886	352	14986	1624	20234	73476
Berlin 4	437	17196	10392	5170	7908	208	22495	1348	5484	70638
Summe	196439	1266375	693367	561694	426150	102663	1231304	157735	831556	5467283

Tabelle A2: Tatsächliche Werte 1994

Bezirk	Land	En/Gew	Bau	Handel	Verkehr	Banken	Dienst	Org./Hh	Körp.	Summe
Neubdb	12683	22809	23280	16725	10510	3191	35837	8322	23136	156493
Rostock	7516	30625	22999	21305	18348	3100	52222	10447	18773	185335
Schwer	9288	32955	26345	18737	13124	3830	40710	9193	23633	177815
Strals	4683	17044	16658	13366	8030	1623	35037	5490	14156	116087
Cottbus	8773	65711	33953	22811	17873	3590	48943	5507	32201	239362
Ebersw	5986	19017	15246	10305	6842	1334	22212	4318	16477	101737
Fkfr/O	7347	26620	22447	15721	10511	2423	34722	8646	22909	151346
Neurup	10817	32536	23702	16852	9784	2118	32124	5837	24581	158351
Potsdam	8653	40155	36635	27127	20394	4839	59025	11363	42968	251159
Dessau	2404	21508	11901	10326	6363	1389	17344	2683	13678	87596
Halbers	4387	21562	12579	10300	5893	1451	19798	1758	16272	94000
Halle	4407	32897	32153	18893	19841	2837	48280	3613	32313	193234
Magdeb	6237	55034	42913	28282	19338	5222	48403	7631	44144	257204
Mersebg	4304	37475	21671	13715	8285	1605	19753	2482	18524	127814
Sangerh	2513	16833	10398	8776	4302	1158	14712	2324	14913	75929
Stendal	6323	12268	13687	8641	6056	1545	12841	2298	16957	80616
Wittenb	2950	12252	8341	4371	2735	818	9816	944	6320	48547
Altenbg	1667	8024	6034	4409	1611	565	5503	788	5849	34450
Annabg	3221	30480	14814	9994	4140	1802	24077	1917	13830	104275
Bautzen	6791	58546	34904	20673	12115	3606	45829	6621	25237	214322
Chemn	3920	50050	25620	26756	16658	5334	49838	7145	25862	211183
Dresden	3449	51943	30092	28895	22842	7012	80790	11756	31636	268415
Leipzig	6630	86676	53151	42590	28604	8988	107467	9075	29780	372961
Oschatz	3539	14321	9208	6363	2284	883	10544	1286	7259	55687
Pirna	6704	31389	21658	10409	4992	1813	30956	2113	9611	119645
Plauen	2485	29524	15411	9424	5610	1443	21950	2683	8618	97148
Riesa	3223	21188	12662	8496	4642	1215	14947	2003	7163	75539
Zwickau	3609	37764	21913	15271	8768	2332	35067	4043	12039	140806
Erfurt	6307	41137	29706	23008	16731	5284	52415	6789	33489	214866
Gera	4473	27121	18089	11970	7202	2286	23133	3460	16224	113958
Gotha	5679	32251	17063	15223	6494	1891	26075	6754	14340	125770
Jena	4275	35506	16432	12640	6546	2146	35905	4063	14933	132446
Nordhn	3800	21807	14612	10227	4943	1356	21831	2308	15440	96324
Suhl	5761	46682	22938	15542	8852	2885	33971	4983	24664	166278
Berlin 1	995	17957	9886	20965	28151	7810	71820	7578	29817	194979
Berlin 2	199	13979	11754	8521	3382	334	33784	3475	18405	93833
Berlin 3	619	12786	11710	8441	2779	951	16441	2216	19049	74992
Berlin 4	582	15588	9787	4901	8261	218	21808	1406	5274	67825
Summe	187199	1182020	782352	580971	393836	102227	1315930	185318	750474	5480327

Tabelle A3: Aggregierte Randsummen und Zusatzinformationen

Bezirk	Land	En/Gew	Bau	Handel	Verkehr	Banken	Dienst	Org./Hh	Körp.	Summe
Neubdb										
Rostock										
Schwer										
Strals					<			>		
Cottbus					115450			69000		
Ebersw										
Fkfr/O										
Neurup							<			
Potsdam							552000			1537658
Dessau										
Halbers										
Halle										
Magdeb										
Mersebg		209829								
Sangerh										
Stendal										
Wittenb										966940
Altenbg										
Annabg										
Bautzen										
Chemn										
Dresden										
Leipzig										
Oschatz										
Pirna										
Plauen										
Riesa										1553625
Zwickau										
Erfurt										
Gera										
Gotha										
Jena										
Nordhn										
Suhl										990448
Berlin 1										
Berlin 2	2395									
Berlin 3										
Berlin 4								<	14700	
Summe	187199	1182020	782352	580971	393836	102227	1315930	185318	750474	5480327

Tabelle A4: Schätzung unter Berücksichtigung aggregierter Randsummen- und Zusatzinformationen

Bezirk	Land	En/Gew	Bau	Handel	Verkehr	Banken	Dienst	Org./Hh	KöRp.	Summe
Neubdb	11173	21358	23760	17840	9767	3098	35418	7716	23754	153884
Rostock	8098	30183	23727	21936	18436	3207	53519	11290	20054	190450
Schwer	10484	31197	26388	19322	13364	3681	41059	10134	24820	180450
Strals	5224	17529	16695	12310	9359	1613	35296	5478	14893	118396
Cottbus	9101	70493	31784	22571	18064	3603	46562	5657	32066	239900
Ebersw	6481	20478	17920	10969	6980	1567	23540	4571	18388	110894
Fkfrt/O	6892	26272	20619	13792	10361	2263	31867	8462	22851	143379
Neurup	11346	31963	23214	16125	8942	2231	30995	5609	22234	152660
Potsdam	8726	38399	37636	24825	20176	4802	58337	10263	44508	247672
Dessau	2627	20733	13020	9864	6343	1475	16678	3621	11720	86082
Halbers	4006	20778	13687	10830	5363	1465	21303	1555	14581	94566
Halle	2284	36014	28187	20002	21546	2887	51806	3910	32769	199404
Magdeb	6608	53658	37415	29082	20600	5356	49661	9371	45964	257713
Mersebg	4805	39472	20450	13178	6947	1657	20303	2107	18339	127258
Sangerh	2780	16234	11303	8590	4020	1581	13346	3148	14293	75295
Stendal	7086	10948	14440	8281	6752	1528	12978	2227	15500	79740
Wittenb	3102	11992	8498	4292	2157	806	9335	852	5849	46882
Altenbg	1706	7818	5953	4094	1594	548	6309	535	4345	32902
Annabg	3315	30905	13677	9335	4256	1708	24534	1701	12969	102399
Bautzen	6766	56978	34776	19706	12119	3709	44624	5149	26870	210697
Chemn	4034	51682	28436	30864	13825	5349	48933	8039	27568	218730
Dresden	3839	51654	30139	30874	22788	6721	81138	12510	33172	272837
Leipzig	6492	85456	53214	42858	29171	8249	111034	9189	30390	376054
Oschatz	3442	14060	8892	6371	2192	863	10846	1071	6794	54532
Pirna	6366	32110	20537	9422	4743	2293	28468	1761	10396	116096
Plauen	2329	29215	15032	8490	5406	1408	20923	2720	9312	94834
Riesa	2835	21299	12776	7352	4767	1173	15252	1664	7426	74544
Zwickau	3317	38218	21715	16182	9004	2389	33280	3520	14894	142518
Erfurt	6496	40238	30153	23982	16854	5396	52024	7151	32727	215019
Gera	3886	26887	19866	12361	6854	2330	24379	4785	15199	116548
Gotha	4762	30838	17536	14149	6838	1848	27007	6062	12987	122028
Jena	4390	39708	17031	11810	6532	2176	34638	3099	14142	133526
Nordhu	3826	21506	15660	9737	4578	1375	22364	2055	14660	95762
Suhl	6181	47131	23776	14645	8993	2918	34357	3639	23407	165046
Berlin 1	293	18413	10569	23476	30599	8095	72931	9420	32667	206462
Berlin 2	1095	13520	11374	8251	3000	336	32455	2660	16105	88795
Berlin 3	549	11628	11680	8160	2552	331	15367	1432	17198	68896
Berlin 4	458	15053	10821	5040	6992	196	23067	1188	4661	67475
Summe	187199	1182020	782352	580971	393836	102227	1315930	185318	750474	5480327

Relative Ähnlichkeit $\text{sim} = 61,57\%$

Anhang 2: Messung der Güte einer Schätzung durch ENTROP oder RAS

Gegeben seien drei Tabellen gleicher Dimension: Die „wahre“ Tabelle B, die approximiert werden soll (hier die Werte 1994), die Referenztable R, bezüglich der die Verbesserung gemessen wird (hier die Werte 1993) und die Testtable T, deren Güte gemessen werden soll. Alle drei Tabellen werden jeweils auf Gesamtsumme 1 normiert, so daß in die Bewertung nur die Strukturunterschiede und nicht die absoluten Größenunterschiede eingehen. Die normierten Tabellen seien entsprechend mit b , r und t bezeichnet.

Sind $F(t;b)$ und $F(r;b)$ die relativen Entropien der normierten Test- bzw. Referenztable bezüglich der Basistable b , so misst der Ausdruck

$$\text{sim}(T|R,B) = 100 * (1 - E(t,b)/E(r,b))$$

auf einer Skala von 0 bis 100 die Ähnlichkeit der Tabelle T zu der Tabelle B relativ zur Tabelle R (in %).

Für $T = R$, also der Approximation der gesuchten durch die Referenztable, hat man den Wert 0%. Für $T = B$, also exakte Approximation der gesuchten Tabelle, hat man Wert 100%.

$\text{sim}(T|R,B)$ wird als die relative Ähnlichkeit (zu B) bezogen auf die Referenztable bezeichnet. Sie liefert keine Information über die Differenzen bei den einzelnen Zellen, sondern ist lediglich ein globaler Indikator für die Annäherung an die gesuchte Tabelle.

Tabelle A5: Schätzung auf der Basis disaggregierter Randsummen

Bezirk	Land	En/Gew	Bau	Handel	Verkehr	Banken	Dienst	Org./Hh	Körp.	Summe
Neubdb	11290	21647	24451	18069	10276	3147	36647	6923	24043	156493
Rostock	7823	29245	23342	21239	18543	3115	52939	9684	19404	185335
Schwer	10269	30650	26323	18970	13630	3626	41181	8814	24352	177815
Strals	5077	17088	16526	11992	9472	1576	35128	4727	14500	116087
Cottbus	9000	69920	32010	22372	18600	3583	47149	4967	31762	239362
Ebersw	5904	18709	16623	10015	6620	1435	21956	3697	16777	101737
Fkfrt/O	7236	27664	22045	14513	11326	2389	34257	7888	24029	151346
Neurup	11686	33018	24349	16646	9589	2311	32687	5129	22936	158351
Potsdam	8777	38737	38550	25025	21129	4856	60080	9165	44839	251159
Dessau	2618	22051	12828	9852	6279	1478	17018	3775	11698	87596
Halbers	3906	21622	13195	10583	6163	1436	21268	1586	14240	94000
Halle	2197	36985	26816	19289	20594	2793	51041	3936	31582	195234
Magdeb	6473	56101	36238	28553	20046	5275	49812	9605	45099	257204
Mersebg	4714	41328	19835	12957	6770	1634	20394	2163	18019	127814
Sangerh	2750	17141	11056	8518	3951	1573	13519	3259	14163	75929
Stendal	7075	11667	14256	8287	6697	1535	13269	2327	15502	80616
Wittenb	3147	12982	8521	4363	2173	822	9694	904	5942	48547
Altenbg	1804	8048	6224	4339	1674	582	6585	592	4602	34450
Annabg	3415	30997	13932	9640	4356	1769	24949	1834	13384	104275
Bautzen	6956	57020	35347	20304	12377	3834	45278	5539	27667	214322
Chemn	3929	49001	27384	30129	13376	5238	47040	8192	26893	211183
Dresden	3808	49867	29552	30688	22449	6702	79419	12981	32950	268415
Leipzig	6506	83366	52726	43047	29040	8312	109825	9636	30504	372961
Oschatz	3551	14119	9069	6588	2247	895	11043	1156	7020	55687
Pirna	6638	32593	21173	9847	4913	2404	29298	1921	10858	119645
Plauen	2413	29471	15401	8818	5564	1467	21399	2950	9665	97148
Riesa	2905	21247	12944	7551	4852	1208	15425	1784	7622	75539
Zwickau	3312	37152	21440	16196	8932	2398	32801	3678	14897	140806
Erfurt	6544	39459	30033	24213	16866	5465	51727	7537	33022	214866
Gera	3833	25817	19373	12220	6716	2311	23734	4938	15016	113958
Gotha	4951	31214	18027	14746	7063	1932	27716	6595	13526	125770
Jena	4406	38787	16897	11877	6511	2195	34305	3253	14214	132446
Nordhn	3886	21262	15725	9912	4619	1405	22418	2184	14913	96324
Suhl	6294	46724	23939	14948	9098	2988	34533	3877	23876	166278
Berlin 1	233	17296	10083	22043	28476	7625	67436	11133	30653	194979
Berlin 2	978	14264	12188	8702	3136	355	33706	3531	16974	93833
Berlin 3	505	12623	12878	8855	2744	360	16421	1956	18651	74992
Berlin 4	390	15137	11051	5066	6965	197	22833	1504	4682	67825
Summe	187199	1182020	782352	580971	393836	102227	1315930	185318	750474	5480327

Relative Ähnlichkeit sim = 63,07 %

Tabelle A6: Schätzung auf der Basis disaggregierter Randsummen und regionaler Zusatzinformationen

Bezirk	Land	En/Gew	Bau	Handel	Verkehr	Banken	Dienst	Org./Hh	Körp.	Summe
Neubdb	11335	21659	24073	18195	9943	3172	36093	7852	24171	156493
Rostock	7860	29283	23000	21404	17955	3142	52178	10991	19522	185335
Schwer	10307	30656	25908	19096	13184	3652	40545	9992	24473	177815
Strals	5109	17136	16307	12103	9185	1592	34674	5373	14609	116087
Cottbus	9063	70168	31610	22596	18051	3621	46575	5650	32027	239362
Ebersw	5933	18738	16383	10095	6412	1448	21646	4197	16883	101737
Fkfrt/O	7259	27656	21687	14603	10950	2405	33712	8938	24137	151346
Neurup	11745	33072	23999	16781	9289	2331	32228	5824	23083	158351
Potsdam	8828	38828	38024	25246	20481	4902	59278	10413	45159	251159
Dessau	2661	21073	13236	10046	6502	1508	16972	3690	11909	87596
Halbers	3965	20631	13593	10775	6372	1463	21178	1548	14474	94000
Halle	2225	35208	27560	19594	21242	2839	50707	3833	32026	195234
Magdeb	6565	53485	37300	29047	20708	5370	49560	9367	45802	257204
Mersebg	4807	39611	20525	13251	7030	1673	20399	2120	18398	127814
Sangerh	2792	16355	11388	8672	4084	1602	13461	3181	14395	75929
Stendal	7133	11057	14587	8380	6877	1553	13123	2255	15650	80616
Wittenb	3200	12409	8794	4451	2250	839	9671	884	6050	48547
Altenbg	1778	8180	6224	4288	1681	576	6621	561	4541	34450
Annabg	3359	31441	13906	9510	4364	1746	25035	1734	13181	104275
Bautzen	6848	57900	35319	20051	12411	3789	45483	5244	27277	214322
Chemn	3874	49834	27404	29800	13435	5185	47327	7768	26556	211183
Dresden	3756	50732	29584	30363	22555	6636	79930	12312	32547	268415
Leipzig	6403	84619	52664	42494	29110	8211	110280	9118	30062	372961
Oschatz	3499	14348	9068	6510	2255	885	11101	1095	6926	55687
Pirna	6529	33062	21134	9714	4922	2373	29400	1817	10694	119645
Plauen	2373	29896	15373	8699	5574	1448	21475	2790	9519	97148
Riesa	2858	21562	12926	7453	4863	1193	15486	1688	7510	75539
Zwickau	3260	37713	21416	15990	8954	2369	32940	3480	14683	140806
Erfurt	6457	40155	30075	23964	16950	5413	52074	7151	32627	214866
Gera	3781	26262	19393	12090	6747	2288	23884	4683	14831	113958
Gotha	4883	31749	18043	14587	7095	1913	27889	6254	13358	125770
Jena	4333	39343	16866	11717	6522	2167	34424	3076	13998	132446
Nordhn	3830	21612	15729	9798	4637	1390	22542	2070	14717	96324
Suhl	6196	47439	23917	14760	9122	2952	34685	3670	23536	166278
Berlin 1	266	17264	9904	22144	29049	7665	68910	9035	30743	194979
Berlin 2	1112	14220	11956	8731	3195	356	34400	2862	17002	93833
Berlin 3	575	12606	12656	8900	2801	362	16790	1588	18715	74992
Berlin 4	442	15059	10819	5073	7082	198	23256	1216	4680	67825
Summe	187199	1182020	782352	580971	393836	102227	1315930	185318	750474	5480327

Relative Ähnlichkeit sim = 67,70 %