Quelle: Gerhard Kleinhenz (Hrsg.) (2002): IAB-Kompendium Arbeitsmarkt- und Berufsforschung. Beiträge zur Arbeitsmarkt- und Berufsforschung, BeitrAB 250, S. 507-516.

Das ENTROP-Verfahren zur Schätzung von Beständen und Strömen in Bildungswesen und Arbeitsmarkt

Uwe Blien, Alexander Reinberg[*]

1 Einführung

Die Bildungsgesamtrechnung des IAB (BGR) wurde entwickelt, um die Übergänge zwischen Bildungswesen und Arbeitsmarkt konsistent zu beschreiben und zu analysieren und um die Bestände einzelner Geburtsjahrgangskohorten in den verschiedenen, dabei relevanten Statusarten nachzeichnen zu können (vgl. „Reinberg/Hummel: Die Bildungsgesamtrechnung des IAB“).

Die meisten BGR-Bestandsdaten entstammen amtlichen deutschen Statistiken; es bestehen jedoch erhebliche Informationslücken hinsichtlich der Übergänge zwischen den einzelnen Statusarten. Übergangsinformationen liegen zwar aus diversen Erhebungen vor; sie sind jedoch häufig inkonsistent und mit Stichprobenfehlern behaftet. Um derartige heterogene und inkonsistente Informationen dennoch für die Schätzung der Matrix der Ströme von Personen in der BGR nutzbar zu machen, wurde ein neuartiges Verfahren entwickelt (vgl. Blien/Graef 1998). Da es auf dem Prinzip der Entropieoptimierung beruht, wurde es „ENTROP“ genannt. Seine Prinzipien und Anwendung (vgl. auch Blien/Tessaring 2001) sind Gegenstand dieses Beitrags.

Die ENTROP-Methode wird inzwischen innerhalb und außerhalb des IAB für andere Problemstellungen als der BGR genutzt, da sie unter sehr allgemeinen Bedingungen anwendbar ist. Sie kann immer dann verwendet werden, wenn Informationen über die Struktur einer zu schätzenden Matrix („Tabelle“) in Form von linearen Gleichungen und/oder Ungleichungen vorliegen. Außerdem kann eine a priori bekannte Referenzmatrix spezifiziert werden. Das Ergebnis - die Struktur der geschätzten Tabelle - entspricht so weit wie möglich dieser Referenzmatrix und berücksichtigt exakt die spezifizierten Randbedingungen. Tatsächliche oder mögliche Anwendungsgebiete von ENTROP sind:

- die Erstellung von Projektionen (z. B. wenn die verbundene Verteilung einiger Variablen nach einer vorangehenden univariaten Verteilung geschätzt werden soll, vgl. die

[*] Gedankt sei Friedrich Graef (Universität Erlangen-Nürnberg) für seine Arbeiten zur Entwicklung und Programmierung des ENTROP-Algorithmus. Die Verantwortung für den vorliegenden Beitrag liegt jedoch allein bei den Autoren.

Vorausschätzungen der regionalen Beschäftigung bei Tassinopoulos 2000, Blien/Tassinopoulos 2001);

- die Disaggregation von Daten, die lediglich als Summen vorliegen;
- die Gewichtung von Zufallsstichproben, d. h. ihre Anpassung an Verteilungen aus offiziellen Statistiken;
- die Schätzung von Tabellen aus heterogenen, inkonsistenten und unvollständigen Daten (z. B. in der Input-Output-Analyse, vgl. Berwert 2000);
- die Berechnung von Übergangsmatrizen in Markov-Ketten.

Die ENTROP-Methode wurde zur Berechnung von Matrizen entwickelt, um die gegebenen Informationen optimal und verzerrungsfrei zu nutzen. Das Ergebnis kann anhand statistischer Kriterien und der formalen Informationstheorie (Shannon/Weaver 1949, Golan et al. 1996) interpretiert werden, da die geschätzte Tabelle unter definierten Bedingungen am meisten wahrscheinlich ist. Das ENTROP-Verfahren ist eine Generalisierung der bekannten RAS-Methode (Deming/Stephan 1940), die wiederum identisch mit dem Iterative Proportional Fitting-Algorithmus (IPF) ist, der in log-linearen Modellen der Statistik verwendet wird.

2 Das Schätzproblem in der BGR

Lässt man die Unterscheidung nach Kohorten zunächst beiseite, so wird die Beziehung zwischen den nach Statusarten differenzierten Beständen S^t und S^{t+1} in der BGR als inflow-outflow-Matrix X^t hergestellt. In dieser quadratischen Matrix wird der Anfangsstatus von Individuen in den Reihen und der Endstatus in den Spalten aufgeführt. Das System ist konsistent, d. h. die Spaltensummen von S^t entsprechen S^{t+1} und die Zeilensummen entsprechen X^t:

$$s_i^t = \sum_j^n x_{ij} \quad \text{alle i,} \quad s_j^{t+1} = \sum_j^n x_{ij} \quad \text{alle j} \qquad (1),$$

wobei x_{ij} (wir unterdrücken t) die (noch unbekannten) Inhalte der Tabelle, s_i^t die Zeilensummen, s_j^{t+1} die Spaltensummen sind; n ist die Anzahl der Statusarten und identisch mit der Zahl der Tabellenzeilen und -spalten, S^t ist der Vektor aller s_i^t und S^{t+1} ist der Vektor aller s_j^{t+1}. Die Summe aller Zeilensummen Σs_i^t entspricht derjenigen aller Spaltensummen Σs_j^{t+1}. In der BGR ist n=23 (Zahl der Bildungs-, Ausbildungs- und Erwerbskonten). Tabelle 1 zeigt die Übergangsmatrix X^t.

Bei der Vervollständigung der BGR um die Bruttoströme und ihre Herkunft bzw. ihr Ziel, d. h. um die Übergänge zwischen den Statusarten, sind amtliche Statistiken im Allgemeinen wenig ergiebig, da sie hierzu keine ausreichenden Datengrundlagen bereitstellen. Um die interne Struktur der Übergangsmatrix X zu berechnen, müssen für die Erstellung der BGR daher teilweise heterogene und unkompatible Informationen einbezogen werden, so z. B. Ergebnisse verschiedener Stichprobenerhebungen. Die verfügbaren Informationsbausteine sind charakterisiert durch unterschiedliche Aggregationsniveaus und Zuverlässigkeitsgrade.

Tabelle 1: Strommatrix X^t

	Status in t_{0+1}				Σ
Status in t_0	x_{11}	x_{12}	.....	x_{1I}	b^r_1
	x_{21}	x_{22}	.....	x_{2I}	b^r_2
	x_{31}	x_{32}	.....	x_{3I}	b^r_3
	…				…
	x_{J1}	x_{J2}	.....	x_{JI}	b^r_J
Σ	b^c_1	b^c_2	.....	b^c_I	b^{rc}

Die Aufgliederung nach alters- und geschlechtsspezifischen Kohorten kompliziert das Schätzproblem zusätzlich. Alle kohortenspezifischen Übergangsmatrizen X^{tc} müssen in eine Matrix X^{tf} integriert werden, die vier Dimensionen aufweist: Anfangsstatus, Endstatus, Geburtsjahrgangskohorte und Geschlecht. Obwohl Übergänge zwischen verschiedenen Geburtskohorten sowie zwischen Männern und Frauen auszuschließen sind, ist eine Aufgliederung der X^{tf} in verschiedene X^{tc} nicht durchführbar, da Informationen über Strömungsdaten zwischen den einzelnen Statuspositionen in sehr unterschiedlichen Formen und Aggregaten vorliegen: viele empirische Studien unterscheiden nicht nach Alter oder gar nach Geschlecht. In diesen Fällen bezieht sich die Information auf mehrere X^{tc}-Matrizen und macht eine simultane Schätzung erforderlich.

3 Die Information über Strommatrizen

Tabelle 1 illustriert die Strommatrix. Für eine möglichst einfache Darlegung des Schätzproblems und des entwickelten Lösungsansatzes bleibt die Unterscheidung nach verschiedenen alters- und geschlechtsspezifischen Kohorten hier zunächst unberücksichtigt. Damit weist die Übergangsmatrix X^t nur zwei Dimensionen auf.

Ausgangspunkt der Schätzung von Strömen sind die gegebenen Bestandsdaten, d. h. die Summen der Zeilen und Spalten von X^t (vgl. (1)). Diese Konsistenzrestriktion ist eine der Voraussetzungen (und der Vorteile) des gewählten Berechnungsverfahrens. Die Zellen von X^t sind zunächst unbekannt (vgl. Tabelle 2).

Viele der erwähnten heterogenen Informationen über die interne Struktur von X^t, die aus Erhebungen oder anderen Quellen stammen, können als lineare Gleichungen oder als

Ungleichungen betrachtet werden. Ein einfaches Beispiel für eine Ungleichung mit einem einzigen x_{ij} wird in (2) aufgeführt. Es ist adäquat, falls die Information über x_{ij} aus einer repräsentativen Stichprobe stammt, da dann zu erwarten ist, daß der Wert mit bestimmten Zufallsfehlern behaftet ist.

$$b^{l} \leq x_{ij} \leq b^{u} \qquad (2)$$

b^{l} und b^{u} sind die aufgrund einer Fehlerschätzung angenommenen unteren und oberen Grenzen des wirklichen (unbekannten) Wertes.

Tabelle 2: Unbekannte Übergangsmatrix X^{t} mit vorgegebenen Beständen

	Status in t_{0+1}				Σ
Status in t_0	?	?	.....	?	b^{r}_{1}
	?	?	.....	?	b^{r}_{2}
	?	?	.....	?	b^{r}_{3}
	…				…
	?	?	.....	?	b^{r}_{J}
Σ	b^{c}_{1}	b^{c}_{2}	.....	b^{c}_{I}	b^{rc}

Falls für X^{t} nur aggregierte Informationen vorliegen (z. B. größere Altersgruppen oder die Summe von Männern und Frauen), müssen diese Summen einbezogen werden:

$$\sum_{i}^{I} \sum_{j}^{J} a_{kij}\, x_{ij} \leq b_{k} \qquad \text{mit} \quad k = 1 \dots K \qquad (3)$$

Ein System (3) enthält Ungleichungen wie (2) und - falls die a_{kij} entweder gleich 0 oder 1 sind - Gleichungen wie (1). Dann können die Zeilen- und Spaltensummen der Tabellen auch als Bedingungen spezifiziert werden, die in den Schätzungen erfüllt werden müssen.

Mit der Verwendung von Ungleichungen können Informationen unterschiedlicher Zuverlässigkeitsgrade einbezogen werden. Wenn einzelne x_{ij} nur aufgrund grober Expertenschätzungen vorliegen, können sie in (2) mit einer relativ großen Bandbreite zwischen b^{l} und b^{u} versehen werden, oder auch nur mit der Definition einer Unter- oder einer Obergrenze. Falls umgekehrt die vorliegenden Daten als „hart“ oder sicher eingeschätzt werden, ist $b_{l} = b_{u}$ und die Ungleichung wird zu einer Gleichung (Beispiel: Daten aus einer Totalerhebung). Die Spezifikation von Ungleichungen für mit Stichprobenfehlern behaftete Werte ist von hoher Bedeutung. Falls die unterschiedlichen Datenquellen inkonsistent sind, gibt es keine Lösungen für die Übergangsmatrix.

Unter bestimmten Bedingungen kann das System (3) exakt gelöst werden. In komplexen Fällen mit einer großen Zahl heterogener Daten - wie dies in der BGR der Fall ist - ist ein Schätzverfahren erforderlich.

In einigen Fällen kann die Schätzung auf zusätzliche Informationen zur Struktur der Matrix zurückgreifen. Dies ist möglich, falls die Übergangsmatrix X^t für ein bestimmtes Jahr bereits vorliegt, während für benachbarte Jahre lediglich unvollständige und heterogene Informationen bekannt sind. Da anzunehmen ist, dass das Übergangsverhalten sich von Jahr zu Jahr nur relativ geringfügig verändert, könnte die unbekannte Matrix in Übereinstimmung mit der Struktur der bekannten Matrix geschätzt werden (vgl. Tabelle 3).

Tabelle 3: Vorgegebene Referenzmatrix U^{t-1} für eine Ausgangsperiode t_{0-1}

	Status in t_0				Σ
Status in t_{0-1}	u_{11}	u_{12}	.....	u_{1I}	b^r_1
	u_{21}	u_{22}	.....	u_{2I}	b^r_2
	u_{31}	u_{32}	.....	u_{3I}	b^r_3
	…				…
	u_{J1}	u_{J2}	.....	u_{JI}	b^r_J
Σ	b^c_1	b^c_2	.....	b^c_I	b^{rc}

Mit der vorgegebenen Referenzmatrix U^{t-1} kann das Schätzproblem als ein Optimierungsproblem umdefiniert werden: Ein gegebenes Abstandsmaß

$$D_u(X) = \left|X^t - U^{t-1}\right|$$

kann minimiert werden, wobei die vorgegebenen Ungleichungen (3) die Restriktionen des Optimierungsprozesses darstellen. Ein solches Vorgehen ist auch dann angemessen, wenn die Referenzmatrix U^{t-1} selbst das Ergebnis einer Schätzung ist (dann ist $U^{t-1}=X^{t-1}$), aber auf detaillierteren Informationen beruht, als für die Matrix X^t verfügbar sind.

4 Entropieoptimierung

Das beschriebene Schätzproblem erfordert die Definition eines Abstandsmaßes $D_u(X)$ zwischen X^t und U^{t-1} (vgl. Blien/Graef 1998). Nach Prüfung der Eigenschaften einiger Abstandsfunktionen wurde die relative Entropie (4) gewählt.

$$E_u(X) = \Sigma_i \Sigma_j x_{ij} \ln (x_{ij} / u_{ij}) \tag{4}$$

Die Minimierung der relativen Entropie kann statistisch begründet werden. Wenn die Referenzmatrix normiert wurde, sodass $\Sigma_I \Sigma_j q_{ij} = 1$ (mit $q_{ij} = u_{ij}/M$ und $M=\Sigma_I \Sigma_j u_{ij}$), kann q_{ij}

interpretiert werden als die Wahrscheinlichkeit eines Objektes, sich im Status ij zu befinden. x_{ij} kann interpretiert werden als eine Schätzung der Anzahl der Objekte innerhalb einer Gesamtzahl von $N = \Sigma_I \Sigma_j x_{ij}$, die den Status ij eines Systems einnehmen. Kann man von fix gegebenen q_{ij} ausgehen, folgt die Wahrscheinlichkeit der Tabelle X^t einer Multinomialverteilung:

$$P(X) = \frac{N!}{x_{11}!x_{12}! \dots x_{mn}!} q_{11}^{x_{11}} q_{12}^{x_{12}} \dots q_{mn}^{x_{mn}}$$

Der Logarithmus von P(X) ist eng verwandt mit der relativen Entropie $E_u(X)$. Unter Verwendung der Näherung $\ln s! \approx s \ln s - s$ kann algebraisch gezeigt werden:

$$\begin{aligned} \ln P(x) &\approx N \ln N - N - \Sigma\, x_{ij} \ln x_{ij} + \Sigma\, x_{ij} + \Sigma\, x_{ij} \ln q_{ij} \\ &= N \ln (N/M) - E_u(X) \end{aligned}$$

Da der Logarithmus eine monotone Funktion ist, ist das Maximum von ln P(X) gleichzeitig X und das Maximum von P(X). Nimmt man eine stochastisch unabhängige Zuordnung der N Objekte unter den gegebenen Restriktionen (3) an, dann ist die Tabelle mit der Minimierung der relativen Entropie asymptotisch gleich einer Tabelle, die die Verteilung der Objekte mit der höchsten Wahrscheinlichkeit repräsentiert (Blien/Graef 1998). Das beschriebene Vorgehen zur Übersetzung eines Schätz- in ein Optimierungsproblem hat eine lange Tradition in den Natur- und Ingenieurwissenschaften.[1] In den letzten Jahren gibt es in der Ökonometrie ein wieder erwachendes Interesse an der Entropieoptimierung (vgl. Golan et al. 1996).

Das Optimierungsproblem kann nun in folgender Weise geschrieben werden:

$$\text{Min}: \quad E_u(X) = \Sigma_i^I \Sigma_j^J x_{ij} \ln (x_{ij}/u_{ij}) \tag{4}$$

mit den Bedingungen:

$$\begin{aligned} &\Sigma_i^I \Sigma_j^J a_{kij} x_{ij} \leq b_k \quad \text{mit} \quad k = 1 \dots\dots K \\ &0 \leq x_{ij} \end{aligned} \tag{5}$$

[1] Ein besonders eindrucksvolles Beispiel ist die Computertomographie in der Medizin, bei der die Verwandtschaft mit dem diskutierten Problem besonders klar wird. In der Computertomographie werden Röntgen- (oder Ultraschall-)bilder eines Organs unter verschiedenen Winkeln aufgenommen. Jedes resultierende Bild kann man sich als eine extrem fein schattierte Tabelle vorstellen, die Grau- oder Farbwerte für die einzelnen Punkte (= Tabellenzellen) enthält und ein Querschnittsbild des betreffenden Organs darstellt. Das endgültige dreidimensionale Bild wird schließlich mit einem entropie-optimierenden Algorithmus erzeugt, der die (unbekannten) Zwischenwerte zwischen den Querschnitten schätzt. Das Resultat zeigt dann das betreffende Organ in allen Details, einschließlich eventueller pathologischer Veränderungen.

Die Nicht-Negativitäts-Bedingungen werden angeführt, weil alle Elemente von X^t gleich oder größer als 0 sein müssen (man beachte, dass definitionsgemäß alle $u_{ij} \geq 0$ sind). Man erhält die allgemeine Struktur der Ergebnistabelle mittels des Kuhn-Tucker-Karush-Theorems. Die Bedingungen für das Minimum lauten für alle ij-Variablen:

$$\begin{aligned} &\ln (x_{ij} / u_{ij}) + 1 + \Sigma_k \, \mu_k \, a_{kij} = 0 \\ &\mu_k \geq 0 \\ &\mu_k \, (\Sigma \; a_{kij} \, x_{ij} - b_k) = 0 \end{aligned} \tag{6}$$

Die μ_k sind die dualen Variablen. Wenn wir (6) nach den x_{ij} auflösen, dann ist

$$x_{ij} = u_{ij} \cdot e^{-1} \cdot e^{-\Sigma_k \mu_k \cdot a_{kij}} \tag{7}$$

(7) lässt einige Eigenschaften[2] der geschätzten Tabelle erkennen:

- Nichtnegativität: alle $x_{ij} \geq 0$
- Nullerhaltung: wenn einige $u_{ij} = 0$ sind, dann bleiben - für diese ij - die korrespondierenden $x_{ij} = 0$.

5 Der ENTROP-Algorithmus

Aufgabe bei der Schätzung einer Tabelle ist es, die relative Entropie (4) unter den Bedingungen (3) zu minimieren. Dies kann im Prinzip durch die Anwendung einer Vielzahl von Methoden erfolgen, die aus der Theorie der nichtlinearen Optimierung bekannt sind, da die relative Entropie die Eigenschaft einer globalen Konvexität aufweist. Um ein solches Problem zu lösen, haben manche Autoren spezielle Anwendungen vorgeschlagen, so etwa Varianten des Newton-Raphson-Algorithmus (Golan et al. 1996), der geometrischen Programmierung und des Simulated Annealing.

Der gewählte ENTROP-Algorithmus (vgl. Blien/Graef 1998) wurde schließlich für seine Anwendung in der BGR ausgestaltet und programmiert. Da der Algorithmus die spezielle Form der Entropie-Funktion ausschöpft, ist er sehr effizient und spart Computer-Laufzeit. Dieser besondere Vorteil gegenüber anderen Methoden (z. B. der stochastischen Optimierung) ist wichtig, weil jede der Übergangsmatrizen in der BGR etwa 30.000 Elemente enthält und jede Schätzung sich auf ungefähr 2.000 Restriktionen (4) gründet.

---

[2] Es gibt eine zunehmende Zahl von Arbeiten über die Schätzeigenschaften dieses Minimum-Informationsprinzips. Eine Eigenschaft soll gesondert angemerkt werden: Eine Schätzung mit diesem Prinzip ist näherungsweise äquivalent einer Optimierung einer gewogenen Quadratsumme, einer modifizierten Chi-Quadrat Funktion (vgl. Golan et al. 1996).

$$Chi^2 = \sum_i \sum_j \frac{(p_{ij} - q_{ij})^2}{q_{ij}}, \quad \text{wobei} \quad p_{ij} = x_{ij}/N$$

Die ENTROP-Methode basiert auf der Berechnung so genannter Entropieprojektionen. Diese Projektionen sind nichtlineare Gleichungen, wie sie von Censor (1981)[3] zur Maximierung der absoluten Entropiefunktion eingeführt wurde. F. Graef entwickelte eine Modifikation des Censor'schen Algorithmus (vgl. Blien/Graef 1991). Seine Anwendung im ENTROP-Verfahren erlaubt die Spezifikation beliebiger Referenztabellen und Schätzungen, die auf der Optimierung der relativen - und nicht der absoluten - Entropie beruhen.

Der Algorithmus ist ein iteratives Verfahren. Seine Startwerte sind:

$$\begin{aligned} x_{ij} &= u_{ij}\, e^{-1} && \text{alle } i, j \quad \text{und} \\ \mu_k &= 0 && \text{alle } k \end{aligned} \tag{8}$$

Jeder Schritt des Iterationsprozesses enthält folgende Operationen:

1. Berechnung der Entropieprojektion der k'ten Restriktion: bestimme $\delta\mu_k$, sodass:

$$\Sigma\,\Sigma\, a_{kij} \cdot x_{ij} \cdot e^{\delta\mu_k \cdot a_{kij}} = b_k \tag{9}$$

2. Korrektur des Vorzeichens: Wenn $\delta\mu_k > \mu_k$, dann setze $\delta\mu_k := \mu_k$
3. Aktualisierung der Werte für x_{ij} und μ_k:

$$\begin{aligned} x_{ij} &:= x_{ij} \cdot e^{\delta\mu_k \cdot a_{kij}} && \text{alle} \quad i, j \\ \mu_k &:= \mu_k - \delta\mu_k \end{aligned} \tag{10}$$

Es kann gezeigt werden, dass der Iterationsprozess gegen die Lösung von (6) konvergiert, falls eine Lösung existiert. Die Prozedur ist relativ unkompliziert.

Die ENTROP-Methode schließt den Spezialfall des Iterative Proportional Fitting-Algorithmus (IPF) ein, der aus log-linearen statistischen Modellen bekannt ist. Wie Blien/Graef (1998) zeigen, sind die Entropieprojektionen dann mit der Vorgehensweise des IPF identisch, wenn nur die Spalten- und Zeilensummen einer Tabelle als Bedingungen des Optimierungsprozesses vorgegeben sind. Dies gilt ebenso für die in Input-Output-Analysen verwendete RAS-Methode, da sie selbst identisch mit dem IPF ist.

6 Die Anwendung von ENTROP im IAB: BGR und Regionalprojektion

Die ENTROP-Methode kann auch angewandt werden, wenn keine Referenzmatrix verfügbar ist. In diesem Fall wird die Berechnung durchgeführt, indem alle u_{ij} in der Referenztabelle auf 1 gesetzt werden. Das Verfahren versucht dann, die Ergebnismatrix so

[3] Die (absolute) Entropie einer Tabelle ist gegeben durch: $Ea(X) = - \Sigma\, \Sigma\, x_{ij} \ln x_{ij}$
Die relative Entropie hat den gleichen Wert (multipliziert mit –1), wenn $u_{ij} = 1$, alle i,j, d. h. wenn keine Vorgabematrix spezifiziert wurde.

gleichmäßig zu besetzen wie möglich. Ob dies ein vernünftiges Vorgehen ist, hängt von der spezifischen Problemstellung ab. In diesem Falle ergibt die Optimierung der relativen und der absoluten Entropie das identische Ergebnis.

In der Bildungsgesamtrechnung des IAB (BGR) wird bei der Anwendung von ENTROP anders vorgegangen: Es gibt Übergänge, die aus institutionellen Gründen ausgeschlossen sind (z. B. der unmittelbare Übergang von der Hauptschule zur Universität). Diese Übergänge u_{ij} werden in der Referenzmatrix auf Null gesetzt. Sodann werden Näherungswerte für andere u_{ij} eingesetzt, die aufgrund einer Schätzung für benachbarte Jahre gewonnen wurden. Sie werden, falls nötig, durch Experteneinschätzungen korrigiert. Dieses Verfahren ist notwendig, weil die zu schätzende Matrix eine große Zahl fragmentarischer Informationen enthält, die als Bedingungen in den Schätzprozess eingegangen sind. Natürlich kann jede Schätzung im Prinzip nur so zuverlässig sein wie die Informationen, auf die sie sich gründet. Zusätzlich werden Implausibilitäten und Widersprüche der Informationsbasis durch das gewählte Verfahren aufgedeckt.

Wie bereits erwähnt, haben die zu schätzenden Matrizen der BGR vier Dimensionen (Jahresanfangsbestände, Jahresendbestände, Alter, Geschlecht). Eine mit der ENTROP-Methode geschätzte Matrix, die nachträglich etwas aggregiert wurde, ist in Abbildung 3 des Beitrags von Reinberg/Hummel („Die Bildungsgesamtrechnung des IAB“) dargestellt.

Da die ENTROP-Methode für Tabellen mit beliebig vielen Dimensionen angewendet werden kann, ist ihr Einsatz für viele Problemstellungen außerhalb der BGR möglich: Das Verfahren kann z. B. genutzt werden, um Daten zu disaggregieren oder um Zufallsstichproben entsprechend den Randverteilungen amtlicher Statistiken zu gewichten. Im Fall der Anwendung von ENTROP in Regionalprojektionen der Beschäftigung (vgl. Tassinopoulos 2000, Blien/Tassinopoulos 2001) werden Randverteilungen nach Wirtschaftszweigen und Regionen aus verschiedenen externen Informationen generiert. Eine nach diesen beiden Variablen differenzierte Matrix für die Vergangenheit dient als Basismatrix einer ENTROP-Schätzung, die die Randverteilungen und weitere Informationen als Restriktionen verwendet. Das Verfahren konnte mit durchaus beachtenswertem Erfolg bei der Gewinnung nach Landkreisen und kreisfreien Städten differenzierten Beschäftigungsprojektionen angewendet werden. Eine Prognoseevaluation zeigte, dass eine Status-quo-Prognose geschlagen werden kann.

Literatur

Berwert, Adrian (2000): Entrop: A Flexible and Hybrid Approach for the Estimation of Regional Input-Output Tables (Paper presented at the 6th World Congress of the Regional Science Association Internat. Lugano).

Blien, Uwe/Graef, Friedrich (1991): Entropieoptimierungsverfahren in der empirischen Wirtschaftsforschung, in: Jahrbücher für Nationalökonomie und Statistik 208/4: 399-413.

Blien, Uwe/Graef, Friedrich (1998): Entropy Optimizing Methods for the Estimation of Tables, in: Balderjahn, I./Mathar, R./Schader, M. (Hrsg.) (1998): Classification, Data Analysis, and Data Highways (Proceedings of the 21st Annual Conference of the Gesellschaft für Klassifikation), Berlin etc.: Springer: 3-15.

Blien, Uwe/Reinberg, Alexander/Tessaring, Manfred (1990): Die Ermittlung der Übergänge zwischen Bildung und Beschäftigung, in: MittAB 23/2, S. 181-204.

Blien, Uwe/Tassinopoulos, Alexandros (2001): Forecasting Regional Employment with the ENTROP Method, in: Regional Studies 35/2: 113-124.

Blien, Uwe/Tessaring, Manfred (2001): Übergänge zwischen Bildungswesen und Arbeitsmarkt, in: Weizsäcker, R. von (Hrsg.): Bildung und Beschäftigung, Berlin: Duncker & Humblot.

Censor, Yair (1981): Row-actions methods for huge and sparse systems and their applications, in: SIAM Review, Vol. 23: 444ff.

Deming, W.E./Stephan, F.F. (1940): On a least squares adjustment of a sampled frequency table when the expected marginal totals are known, in: The Annals of Mathematical Statistics 11/1, March: 427 ff.

Golan, A./Judge, G./Miller, D. (1996): Maximum Entropy Econometrics. Robust Estimation with Limited Data, New York etc.: John Wiley and Sons.

Reinberg, Alexander/Hummel, Markus (1999): Bildung und Beschäftigung im vereinigten Deutschland, BeitrAB 226, Nürnberg.

Shannon, C.E./Weaver, W. (1949): The Mathematical Theory of Communication, Urbana: University of Illinois Press.

Tassinopoulos, Alexandros (2000): Die Prognose der regionalen Beschäftigungsentwicklung, BeitrAB 239, Nürnberg.