# Probleme der Schätzung von Matrixelementen aus bekannten Randverteilungen

## Dargestellt am Beispiel einer Wanderungsmatrix auf Länderebene für die Bundesrepublik Deutschland

Von Herwig Birg

### Problemstellung

In der empirischen Wirtschaftsforschung steht man häufig vor dem Problem, die Elemente einer Matrix zu schätzen, über die nur spärliche Informationen vorhanden sind. So kennt man manchmal die Summe der Matrixelemente für jede Zeile und Spalte, nicht aber die Elemente selbst. Beispiele hierfür sind Wanderungsmatrizen, Verkehrsstrommatrizen und Input-Output-Tabellen.

Liegt eine vollständig ausgefüllte Matrix vor, die den betrachteten Sachverhalt lediglich für einen anderen als den gewünschten Zeitpunkt beschreibt, so versucht man in der Regel, aus der Struktur dieser bekannten vollständigen Matrix auf die Elemente der unbekannten Matrix zu schließen. Eines der dabei angewandten Schätzverfahren ist das der doppelten Proportionalität[1]. Dieses Verfahren beruht auf der Hypothese, daß der Wachstumsfaktor eines zu schätzenden Matrixelementes $X_{ij}$ gleich dem geometrischen Mittel aus dem Wachstumsfaktor der Zeilensumme $X_{i.}$ und dem Wachstumsfaktor der Spaltensumme $X_{.j}$ ist. Es wird also zwischen dem bekannten Element $X^0_{ij}$ zum Zeitpunkt $t_0$ und dem unbekannten Element $X^1_{ij}$ zum Zeitpunkt $t_1$ folgender Zusammenhang postuliert

$$X^1_{ij} = \sqrt{\frac{X^1_{i.}}{X^0_{i.}} \frac{X^1_{.j}}{X^0_{.j}}}\, X^0_{ij} \tag{1}$$

Werden die unbekannten Matrixelemente nach Formel (1) berechnet, so zeigt sich in der Regel, daß die geschätzten Elemente mit den vorgegebenen Spalten- und Zeilensummen inkonsistent sind. Um diese Inkonsistenz zu beseitigen, können mehrere Wege beschritten werden. Es liegt nahe, jedes geschätzte Element um eine additive Restgröße $e_{ij}$ zu korrigieren, wobei die $e_{ij}$ so bestimmt werden, daß der quadratische Ausdruck $\sum_i \sum_j e^2_{ij}$ ein Minimum annimmt und gleichzeitig die Konsistenzbedingungen erfüllt sind. Ein anderer, von der Rechenzeit her weniger aufwendiger Ansatz liegt dem MODOP-Verfahren zugrunde. Bei diesem Verfahren werden die geschätzten Elemente jeder Zeile mit einem Faktor multipliziert, der zur vorgegebenen Zeilensumme führt. Ist auf diese Weise für jede Zeile die Konsistenz erreicht, wird mit einem entsprechenden Proportionalitätsansatz die Konsistenz der Spalten hergestellt, dann wieder die der Zeilen usf. Das Iterationsverfahren konvergiert in den meisten Fällen[2] nach einer relativ geringen Anzahl von Schritten zu den spalten- und zeilenweise konsistenten Elementen[3].

Wichtig bei diesem Verfahren ist die Tatsache, daß die Anfangswerte der Iteration frei wählbar sind. Allerdings sind die Werte, gegen die das Verfahren konvergiert, von diesen Anfangswerten nicht unabhängig, obwohl in jedem Falle zeilen- und spaltenweise Konsistenz erreicht wird. Die Frage, mit welchen Anfangswerten der Iterationsprozeß gestartet werden soll, verdient deshalb einige Beachtung. Sie ist nicht für alle Verfahren, die zur Schätzung von Matrizen verwendet werden, relevant. Beim RAS-Verfahren beispielsweise wird so vorgegangen, daß man für jede Zeile $i$ einen Faktor $r_i$ und für jede Spalte $j$ einen Faktor $s_j$ sucht, für die das Produkt dieser beiden Faktoren mit dem Element $X_{ij}$ der bekannten Matrix das entsprechende Element der gesuchten Matrix ergibt[4]. Die Multiplikatoren $r_i$ und $s_j$, die den

[1] Vgl. R. Stäglin: MODOP — Ein Verfahren zur Erstellung empirischer Transaktionsmatrizen. In: Anwendung statistischer und mathematischer Methoden auf sozialwissenschaftliche Probleme. Arbeiten zur angewandten Statistik, Physika Verlag, Würzburg 1972 (in Vorbereitung).

[2] Zu den mathematischen Konvergenzbedingungen vgl. H. J. Jacksch u. K. Konrad: „Zur Schätzung von Input-Output-Tabellen aus ihren Reihensummen". In: Jahrbücher für Nationalökonomie und Statistik, Bd. 186, Heft 2 (Dez. 1971), S. 131 - 138.

[3] Beim MODOP-Verfahren, das im DIW programmiert wurde, besteht die Möglichkeit, bei der Iteration einzelne Matrixelemente vorzugeben, über deren numerische Werte man feste Anhaltspunkte besitzt.

[4] Vgl. z. B.: M. Bacharach: Biproportional Matrices and Input-Output-Change, Cambridge 1970, S. 19 f.

in der Input-Output-Technologie wirkenden Substitutions- und Fortschrittseffekt beschreiben sollen, sind dabei so zu wählen, daß zeilen- und spaltenweise Konsistenz der geschätzten Elemente mit den vorgegebenen Randverteilungen erreicht wird. Da die Größen $r_i$ und $s_j$ multiplikativ verknüpft sind, ergibt sich zur Bestimmung der gesuchten Werte ein nichtlineares Gleichungssystem, für dessen Lösung ebenso wie beim MODOP-Verfahren ein Algorithmus angewandt werden muß. Die Frage, welches die günstigsten Ausgangswerte für die Iteration sind, hat hier jedoch eine andere Bedeutung; beim MODOP-Verfahren sind es die Matrixelemente, für die geeignete Ausgangswerte gefunden werden müssen, beim RAS-Verfahren dagegen die Faktoren $r_i$ und $s_j$.

Im folgenden wird die Frage diskutiert, wie man beim MODOP-Verfahren zu geeigneten Anfangswerten für die Iteration gelangen kann. Dabei wird versucht, die Formel (1), die zur Bildung geeigneter Anfangswerte dient, und zunächst nicht mehr als eine plausible a priori-Annahme zu sein scheint, zu begründen bzw. zu relativieren. Dabei wird von einem wahrscheinlichkeitstheoretischen Ansatz zur Ermittlung von Matrixelementen ausgegangen. Das Ergebnis der theoretischen Erörterungen wird durch ein empirisches Beispiel ergänzt.

### Ein wahrscheinlichkeitstheoretischer Ansatz[5]

Es sei angenommen, daß zum Zeitpunkt $t_0$ eine vollständige Matrix $\left[X_{ij}^0\right]$ existiert, während für einen anderen Zeitpunkt $t_1$ nur die Zeilen- und Spaltensummen $X_{i.}^1$ und $X_{.j}^1$ bekannt sind. Die unbekannten Elemente der Matrix $\left[X_{ij}^1\right]$ sollen geschätzt werden. Da es sehr viele Möglichkeiten gibt, die Matrix $\left[X_{ij}^1\right]$ so auszufüllen, daß die resultierenden Zeilen- und Spaltensummen mit den vorgegebenen Werten übereinstimmen, wird gefragt, welche dieser Möglichkeiten die größte Wahrscheinlichkeit besitzt, realisiert zu werden. Handelt es sich z. B. um eine Wanderungsmatrix, so ist durch die Zeilen- und Spaltensummen die Gesamtsumme der wandernden Personen gegeben. Die Anzahl der Möglichkeiten, aus der Gesamtsumme $Z^1$ der Matrixelemente das erste Element $X_{11}^1$ zu bestimmen, multipliziert mit der Anzahl der Möglichkeiten, aus dem Rest $Z^1 - X_{11}^1$ das Element $X_{12}$ auszuwählen – usf. bis zum letzten Element, ergibt die Gesamtzahl $N$ der Zuordnungsmöglichkeiten von wandernden Personen zu Matrixfeldern. Die Zahl $N$ ist dabei eine Funktion der einzelnen Elemente $X_{11}^1$, $X_{12}^1, \ldots$, für die man sich bei der Auswahl entschieden hat

$$(2) \qquad N = \frac{Z^1!}{X_{11}^1!\left(Z^1-X_{11}^1\right)!} \cdot \frac{\left(Z^1-X_{11}^1\right)!}{X_{12}!\left(Z^1-X_{11}^1-X_{12}^1\right)!} \cdots = \frac{Z^1!}{\prod\limits_{ij} X_{ij}^1!}\,.$$

Da diejenige Auswahl der Elemente $X_{11}^1, X_{12}^1, \ldots$, für die $N$ am größten ist, die höchste Wahrscheinlichkeit besitzt, findet man die wahrscheinlichste Matrix dadurch, daß man $N$ in bezug auf die Unbekannten $X_{11}^1, X_{12}^1, \ldots$ maximiert[6]. Da aber ferner gefordert wird, daß die Summe der Elemente den vorgegebenen Zeilen- und Spaltensummen entspricht, sind bei der Maximierung folgende Nebenbedingungen zu berücksichtigen

$$(3) \qquad \sum_j X_{ij}^1 = X_{i.}^1 \qquad \text{für alle } i\,,$$

$$(4) \qquad \sum_i X_{ij}^1 = X_{.j}^1 \qquad \text{für alle } j\,.$$

Statt $N$ zu maximieren, genügt es, die monotone Transformation $\ln N$ zu maximieren. Man erhält dann aus (2)

$$(5) \qquad \ln N = \ln Z^1! - \sum_{ij} \ln X_{ij}^1!$$

Da sowohl (2) als auch (5) diskrete Funktionen sind, können die zur Maximierung benötigten partiellen Ableitungen nur gebildet werden, wenn (5) in eine stetige Funktion umgewandelt wird. Hierfür kann die Stirlingsche Näherungsformel verwendet werden. Man erhält dann die Approximation

$$(6) \qquad \ln X_{ij}^1! \approx X_{ij}^1 \ln X_{ij}^1 - X_{ij}^1$$

Setzt man (6) in (5) ein, so ergibt sich

$$(7) \qquad \ln N = \ln Z^1! - \sum_{ij}\left(X_{ij}^1 \ln X_{ij}^1 - X_{ij}^1\right)$$

Um das Maximum von (7) zu finden, empfiehlt es sich, in (7) nicht nach den Absolutwerten $X_{11}^1$, $X_{12}^1, \ldots$, der unbekannten Elemente, sondern nach den Anteilen dieser Elemente an $Z^1$ zu differenzieren. Setzt man die Definitionsgleichung

$$p_{ij}^1 = \frac{X_{ij}^1}{Z^1}$$

---

[5] Vgl. A. G. Wilson: Entropy in Urban and Regional Modelling, London 1970, S. 5 ff.

[6] Auf die implizite Annahme stochastischer Unabhängigkeit der Wanderungen wird weiter unten eingegangen.

in (7) ein, so erhält man

$$(8) \qquad \ln N = \ln Z^1! - \sum_{ij}\left(Z^1 p^1_{ij} \ln Z^1 p^1_{ij} - Z^1 p^1_{ij}\right)$$
$$= \ln Z^1! - \sum_{ij}\left[Z^1 p^1_{ij}\left(\ln Z^1 + \ln p^1_{ij}\right) - Z^1 p^1_{ij}\right]$$
$$= \ln Z^1! - Z^1 \sum_{ij}\left[p^1_{ij}\left(\ln Z^1 + \ln p^1_{ij}\right) - p^1_{ij}\right]$$
$$= \ln Z^1! - Z^1 \sum_{ij}\left[p^1_{ij} \ln Z^1 + p^1_{ij} \ln p^1_{ij} - p^1_{ij}\right]$$
$$= \ln Z^1! - Z^1 \sum_{ij}\left[p^1_{ij}\left(\ln Z^1 - 1\right) + p^1_{ij} \ln p^1_{ij}\right]$$
$$= \ln Z^1! - Z^1 \left\{\left[\sum_{ij} p^1_{ij}\left(\ln Z^1 - 1\right)\right] + \sum_{ij} p^1_{ij} \ln p^1_{ij}\right\}$$
$$= \ln Z^1! - Z^1 \left\{\left(\ln Z^1 - 1\right) + \sum_{ij} p^1_{ij} \ln p^1_{ij}\right\}$$
$$= \ln Z^1! - Z^1 \ln Z^1 + Z^1 - Z^1 \sum_{ij} p^1_{ij} \ln p^1_{ij}$$

Da in (8) $Z^1$ und damit auch $\ln Z^1$ sowie $\ln Z^1!$ Konstante sind, genügt es, die Funktion

$$(9) \qquad \psi\left(p_{11}, p_{12}, \ldots\right) = -\sum_{ij} p^1_{ij} \ln p^1_{ij}$$

in bezug auf die Nebenbedingungen (3) und (4) zu maximieren. Der Ausdruck (9) ist dabei gleich der Entropie der Matrix $\left[p^1_{ij}\right]$. Für das Maximum der Entropie einer Matrix gilt folgender wichtiger Satz, der sich zur Lösung von (9) anwenden läßt: Die Entropie einer Matrix ist kleiner oder gleich der Entropie der Zeilensummen zuzüglich der Entropie der Spaltensummen. Gleichheit besteht genau dann, wenn die Elemente $p^1_{ij}$ nach der Formel

$$(10) \qquad p^1_{ij} = p^1_{.i}\, p^1_{.j}$$

bestimmt werden[7]. Berechnet man also die unbekannten Elemente $p^1_{11}, p^1_{12}, \ldots$ nach Formel (10), so nimmt (9) den höchsten Wert an, den dieser Ausdruck überhaupt besitzen kann. Deshalb erfüllen die nach Formel (10) bestimmten Elemente die geforderte Maximum-Bedingung für die Zielfunktion (9). Darüber hinaus sind die Nebenbedingungen (3) und (4) bei der Bestimmung der Elemente nach Formel (10) automatisch erfüllt. Das Problem, die wahrscheinlichste Matrix $p^1_{ij}$ zu finden, die die Nebenbedingungen (3) und (4) erfüllt, läßt sich also durch Anwendung von Formel (10) lösen.

### Modifikation des Ansatzes

Die nach Formel (10) bestimmten Elemente werden vermutlich von den wahren Elementen mehr oder weniger stark abweichen, und zwar vor allem deshalb, weil in Formel (10) implizit stochastische Unabhängigkeit vorausgesetzt wurde; denn nur dann ist die Entropie der Matrix gleich der Summe der Entropien der Randvektoren. Im Falle einer Wanderungsmatrix für Bundesländer ist diese Voraussetzung nicht gegeben, denn die Annahme stochastischer Unabhängigkeit würde bedeuten, daß die Verteilung einer bestimmten Anzahl von Zuzügen auf die einzelnen Länder unabhängig davon ist, aus welchem Land die Zuzüge stammen. Beispielsweise wäre die Aufteilung der Zuzüge auf Schleswig-Holstein und Bayern unabhängig davon, ob die gewanderten Personen aus Hamburg stammen, das Schleswig Holstein benachbart ist, oder aus Südwürttemberg, das an Bayern grenzt.

Das Beispiel zeigt, daß die Annahme stochastischer Unabhängigkeit oft nicht erfüllt ist, und daß deshalb die nach dem wahrscheinlichkeitstheoretischen Ansatz berechneten Matrixelemente von den wahren Elementen abweichen werden. Die Gründe für diese Abweichungen weisen zugleich den Weg, wie man von der hypothetischen Schätzung zu einer wirklichkeitsnahen Schätzung kommen kann.

Im Falle einer Wanderungsmatrix beruhen die Abweichungen zwischen den hypothetischen und den wahren Werten vor allem darauf, daß beispielsweise ein Hamburger weniger leicht nach Bayern und dafür eher nach Schleswig-Holstein zieht, als ein Württemberger. Dies mag z. B. darauf zurückzuführen sein, daß (a) die Entfernung zu Bayern größer ist als zu Schleswig-Holstein, (b) die landsmannschaftlichen Eigentümlichkeiten von Hamburg und Schleswig-Holstein mehr Verwandtschaft zeigen als die von Hamburg und Bayern. Im Falle einer Input-Output-Tabelle sind es die technologischen und ökonomischen Gegebenheiten der Güterproduktion, die zu bestimmten Verflechtungsstrukturen führen, und die eine Lieferbeziehung beispielsweise zwischen der Stahlindustrie und der Landwirtschaft von vornherein unwahrscheinlicher erscheinen lassen als zwischen der Stahlindustrie und dem Bergbau. Derartige rigide Strukturen lassen sich dadurch feststellen, daß man in der bekannten vollständigen Matrix für den Zeitpunkt $t_0$ die relativen Abweichungen zwischen den nach Formel (10) errechneten hypothetischen Elementen und den tatsächlichen mißt. Man erhält dann für jedes Element der Matrix $p^0_{ij}$ eine Zahl $k^0_{ij}$

$$(11) \qquad k^0_{ij} = \frac{p^0_{i.}\, p^0_{.j} - \dfrac{X^0_{ij}}{\sum\limits_{ij} X^0_{ij}}}{\dfrac{X^0_{ij}}{\sum\limits_{ij} X^0_{ij}}},$$

die die Differenz zwischen dem hypothetischen Element $\left(p^0_{i.}\, p^0_{.j}\right)$ und dem tatsächlichen $\left(X^0_{ij} / \sum_{ij} X^0_{ij}\right)$ im Verhältnis zur Größe des tatsächlichen Elements ausdrückt.

[7] Vgl. z. B. N. W. Smirnow u. J. W. Dunin-Barowski: Mathematische Statistik in der Technik, Deutscher Verlag der Wissenschaften, Berlin (Ost), 1963, S. 47.

Geht man davon aus, daß die Ursachen, die im Basisjahr $t_0$ zu einer Abweichung der tatsächlichen Elemente von den hypothetischen geführt haben, auch in künftigen Perioden derartige Abweichungen verursachen werden, so liegt der Gedanke nahe, die $k_{ij}^0$ als Korrekturfaktoren zu verwenden, die im Zeitablauf mehr oder minder konstant sind.

Verwendet man als erste vorläufige Hypothese die Annahme, daß sich die $k_{ij}$ im Zeitablauf nicht ändern:

$$(12) \qquad k_{ij}^0 = k_{ij}^1 = \frac{p_{i.}^1\, p_{.j}^1 - \dfrac{X_{ij}^1}{\sum\limits_{ij} X_{ij}^1}}{\dfrac{X_{ij}^1}{\sum\limits_{ij} X_{ij}^1}} = \text{konstant,}$$

so kann die Hypothese (12) zur Berechnung der „tatsächlichen Elemente" im Zeitpunkt $t_1$ herangezogen werden: Durch Auflösen von (12) nach $X_{ij}^1 \,/\, \sum\limits_{ij} X_{ij}^1$ bzw. nach $p_{ij}^1$ ergibt sich

$$(13) \qquad p_{ij}^1 = \frac{p_{i.}^1\, p_{.j}^1}{k_{ij}^0 + 1}$$

Gleichung (13) kann umgeformt werden zu

$$(13.1) \qquad p_{ij}^1 = \frac{p_{i.}^1\, p_{.j}^1}{\dfrac{p_{i.}^0\, p_{.j}^0 - \dfrac{X_{ij}^0}{\sum\limits_{ij} X_{ij}^0}}{\dfrac{X_{ij}^0}{\sum\limits_{ij} X_{ij}^0}} + 1}$$

$$= \frac{p_{i.}^1\, p_{.j}^1}{p_{i.}^0\, p_{.j}^0} \cdot p_{ij}^0$$

Aus Gleichung (13.1) ergibt sich folgende Regel zur Berechnung der unbekannten Elemente: Das Element $p_{ij}^1$ ist gleich dem Produkt aus dem bekannten Element $p_{ij}^0$ und den beiden bekannten Wachstumsfaktoren der betreffenden Zeilen- bzw. Spaltensumme[8].

Die Hypothese, daß die Korrekturfaktoren $k_{ij}^0$ im Zeitablauf konstant sind, wurde am Beispiel zweier Wanderungsmatrizen für die Jahre 1959 und 1969 überprüft, die die Wanderungen zwischen den Ländern der BRD beschreiben. Die Faktoren $k_{ij}^0$ und $k_{ij}^1$ wurden zu diesem Zweck paarweise als Punkte einer Geraden aufgefaßt (vgl. Grafik). Damit die Hypothese (12) zutrifft, müssen drei Bedingungen erfüllt sein: (a) es muß der Korrelationskoeffizient für die Wertepaare $(k_{ij}^1 / k_{ij}^0)$ groß sein, (b) es muß sich für die Gerade eine Steigung von + 1 ergeben und schließlich (c) muß das Absolutglied der Geraden Null sein.

Es wurde folgende Gerade ermittelt

$$(14) \qquad k_{ij}^{1969} = 0{,}53 + 0{,}75\ k_{ij}^{1959}\ ;$$

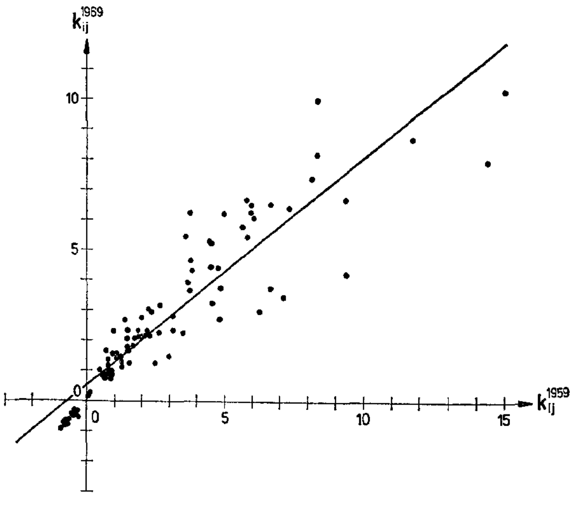
Korrelationskoeffizient = 0,92.

Während also die errechnete Anpassung recht gut ist, sind die Bedingungen (b) und (c) nicht ganz erfüllt.

Da die Steigung der Geraden kleiner als 1 ist, läßt sich folgern, daß die Abweichungen zwischen den hypothetischen und den tatsächlichen Wanderungen in den Jahren von 1959 bis 1969 geringer geworden sind. Dies legt den Gedanken nahe, für jedes Jahr des Beobachtungszeitraumes die Korrekturfaktoren zu ermitteln und zu überprüfen, ob sich für die einzelnen Faktoren extrapolierbare Trends ermitteln lassen. Aus einer Extrapolation dieser Trends ließe sich dann zusammen mit einer Prognose der Randverteilungen die Wanderungsmatrix beispielsweise für 1980 prognostizieren. Weiter unten wird ein ähnliches Prognoseverfahren geschildert.

**Vergleich des modifizierten Ansatzes mit dem Verfahren der doppelten Proportionalität**

Das Verfahren der doppelten Proportionalität zur Schätzung von Matrixelementen beruht auf der in (1) ausgedrückten Hypothese. Vergleicht man diese Hypothese mit der Hypothese (12) in der Form von (13.1), so zeigt sich, daß (1) und (13.1) bis auf das Wurzelzeichen identisch sind[9].




[8] Zu dem gleichen Resultat gelangt H. Theil auf einem anderen Weg, indem er den Logarithmus des Verhältnisses des tatsächlichen Elements zum hypothetischen im Zeitablauf konstant hält. Vgl. H. Theil: Economics and Information Theory, Amsterdam 1967, S. 358, Gleichung (1.4).

[9] Formel (1) gilt sowohl für die Absolutwerte der Elemente $X_{ij}$ als auch für deren Anteile an der Gesamtsumme aller Elemente der Matrix.

Es läßt sich nun die Frage formulieren, welche Funktion $\varphi\left(k_{ij}^0\right)$ im Verfahren der doppelten Proportionalität implizit enthalten ist, wenn (13.1) mit (1) völlig übereinstimmen soll. Als Lösung erhält man

$$k_{ij}^1 = \varphi\left(k_{ij}^0\right) = \sqrt{\frac{p_{i.}^1\, p_{.j}^1}{p_{ij}^0}}\ \sqrt{k_{ij}^0 + 1} - 1 \tag{15}$$

Ersetzt man $k_{ij}^0$ in (13) durch (15), so steht über dem Bruchstrich in (13.1) ein Wurzelzeichen, und die beiden Verfahren sind identisch. Dem modifizierten Ansatz liegt die Hypothese zugrunde, daß sich die $k_{ij}$ im Zeitablauf nicht ändern. Eine sinnvolle Interpretation der Hypothese (15), die dem Verfahren der doppelten Proportionalität entspricht, dürfte dagegen äußerst schwierig sein. Daher erscheint die Frage berechtigt, ob (1) nicht so modifiziert werden sollte, daß das Wurzelzeichen verschwindet. In diesem Falle wäre das modifizierte wahrscheinlichkeitstheoretische Verfahren identisch mit dem Verfahren der doppelten Proportionalität, und zwar unter Zugrundelegung der Hypothese (12) (zeitliche Konstanz der relativen Abweichungen zwischen den hypothetischen und den wahren Matrixelementen). Statt der Hypothese (12) können jedoch auch andere Annahmen gesetzt werden, die zu einer anderen Korrektur von (1) führen würden als zu derjenigen, die darin besteht, das Wurzelzeichen wegzulassen.

**Geschätzte und tatsächliche Wanderungen 1959 und 1969 in der Bundesrepublik Deutschland nach Ländern**

| von \ nach | Schleswig-Holstein | Hamburg | Niedersachsen | Bremen | Nordrhein-Westfalen | Hessen | Rheinland-Pfalz | Baden-Württemberg | Bayern | Saarland | Summe |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Tatsächliche Wanderungen im Jahre 1959[1]) [2]) | | | | | | |
| Schleswig-Holstein | 90 047 | 14 939 | 9 124 | 2 465 | 9 456 | 1 939 | 1 229 | 3 619 | 3 175 | 169 | 136 162 |
| Hamburg | 15 954 | – | 9 644 | 1 164 | 6 330 | 2 433 | 838 | 3 186 | 2 880 | 187 | 42 616 |
| Niedersachsen | 10 996 | 11 372 | 215 894 | 12 630 | 43 487 | 9 718 | 3 793 | 10 696 | 8 346 | 696 | 327 628 |
| Bremen | 1 528 | 1 253 | 7 858 | 948 | 2 619 | 851 | 590 | 1 319 | 847 | 75 | 17 888 |
| Nordrhein-Westfalen | 10 768 | 7 168 | 39 117 | 3 450 | 495 641 | 22 132 | 21 987 | 24 861 | 22 841 | 2 768 | 650 733 |
| Hessen | 2 355 | 2 175 | 8 332 | 872 | 19 253 | 131 406 | 11 773 | 13 890 | 12 716 | 1 122 | 203 894 |
| Rheinland-Pfalz | 1 447 | 1 014 | 3 555 | 710 | 24 887 | 14 226 | 97 973 | 13 985 | 7 246 | 8 081 | 173 124 |
| Baden-Württemberg | 3 516 | 3 104 | 7 737 | 1 024 | 19 511 | 13 711 | 11 145 | 340 052 | 39 132 | 1 945 | 440 877 |
| Bayern | 2 941 | 2 564 | 6 372 | 839 | 18 236 | 12 746 | 6 389 | 42 710 | 392 507 | 936 | 486 240 |
| Saarland | 198 | 108 | 361 | 53 | 1 988 | 1 089 | 4 518 | 2 064 | 850 | 34 216 | 45 445 |
| Summe | 139 750 | 43 697 | 307 994 | 24 155 | 641 408 | 210 251 | 160 235 | 456 382 | 490 540 | 50 195 | 2 524 607 |
| | | | | | *Tatsächliche Wanderungen im Jahre 1969[2])* | | | | | | |
| Schleswig-Holstein | 128 309 | 20 246 | 13 755 | 2 122 | 11 395 | 3 964 | 1 660 | 5 392 | 4 875 | 269 | 191 987 |
| Hamburg | 30 631 | – | 14 893 | 1 359 | 6 859 | 3 180 | 1 001 | 3 910 | 3 955 | 164 | 65 952 |
| Niedersachsen | 14 215 | 13 371 | 327 012 | 14 503 | 44 491 | 13 445 | 4 444 | 13 281 | 11 432 | 659 | 456 853 |
| Bremen | 2 651 | 1 433 | 16 411 | 1 004 | 3 481 | 1 455 | 557 | 1 625 | 1 488 | 69 | 30 174 |
| Nordrhein-Westfalen | 13 026 | 7 169 | 45 271 | 3 543 | 699 842 | 31 931 | 27 883 | 34 722 | 30 963 | 2 894 | 897 244 |
| Hessen | 3 403 | 2 848 | 11 356 | 1 088 | 26 724 | 228 307 | 16 041 | 21 320 | 19 671 | 1 609 | 332 367 |
| Rheinland-Pfalz | 1 845 | 1 039 | 3 855 | 450 | 26 144 | 17 681 | 131 786 | 18 536 | 8 576 | 6 346 | 216 258 |
| Baden-Württemberg | 5 002 | 3 521 | 10 741 | 1 436 | 28 802 | 22 794 | 16 938 | 485 735 | 48 345 | 3 045 | 626 359 |
| Bayern | 4 472 | 3 108 | 9 081 | 1 162 | 24 285 | 18 993 | 7 505 | 48 398 | 531 965 | 1 337 | 650 306 |
| Saarland | 376 | 241 | 907 | 112 | 4 214 | 2 743 | 8 048 | 4 581 | 2 266 | 39 414 | 62 902 |
| *Summe* | 203 930 | 52 976 | 453 282 | 26 779 | 876 237 | 344 493 | 215 863 | 637 500 | 663 536 | 55 806 | 3 530 402 |
| | | | | | Geschätze Wanderungen im Jahre 1969 unter Anwendung des MODOP-Verfahrens | | | | | | |
| Schleswig-Holstein | 130 201 | 18 138 | 13 655 | 2 752 | 12 965 | 3 026 | 1 734 | 4 946 | 4 415 | 155 | 191 987 |
| Hamburg | 24 945 | – | 15 607 | 1 405 | 9 385 | 4 106 | 1 279 | 4 708 | 4 331 | 185 | 65 951 |
| Niedersachsen | 15 327 | 13 311 | 311 467 | 13 592 | 57 480 | 14 622 | 5 160 | 14 032 | 11 188 | 615 | 456 854 |
| Bremen | 2 630 | 1 811 | 13 998 | 1 260 | 4 274 | 1 561 | 991 | 2 146 | 1 402 | 82 | 30 175 |
| Nordrhein-Westfalen | 15 520 | 8 676 | 58 355 | 3 839 | 677 431 | 34 433 | 30 931 | 33 870 | 31 661 | 2 528 | 897 244 |
| Hessen | 3 707 | 2 875 | 13 575 | 1 060 | 28 740 | 223 284 | 18 089 | 20 667 | 19 251 | 1 119 | 332 367 |
| Rheinland-Pfalz | 1 880 | 1 107 | 4 782 | 712 | 30 668 | 19 955 | 124 266 | 17 178 | 9 056 | 6 654 | 216 258 |
| Baden-Württemberg | 5 251 | 3 893 | 11 960 | 1 181 | 27 633 | 22 104 | 16 246 | 480 044 | 56 207 | 1 841 | 626 360 |
| Bayern | 4 086 | 2 991 | 9 163 | 900 | 24 025 | 19 115 | 8 664 | 56 088 | 524 450 | 824 | 650 306 |
| Saarland | 382 | 175 | 720 | 79 | 3 635 | 2 267 | 8 503 | 3 762 | 1 576 | 41 804 | 62 903 |
| Summe | 203 929 | 52 977 | 453 282 | 26 780 | 876 236 | 344 493 | 215 883 | 637 501 | 663 537 | 55 807 | 3 530 405 |
| | | | | | Geschätzte Wanderungen im Jahre 1969 unter Anwendung des modifizierten wahrscheinlichkeitstheoretischen Ansatzes | | | | | | |
| Schleswig-Holstein | 131 439 | 17 770 | 13 557 | 2 682 | 12 635 | 3 062 | 1 689 | 4 775 | 4 214 | 164 | 191 987 |
| Hamburg | 24 480 | – | 15 951 | 1 408 | 9 436 | 4 271 | 1 278 | 4 675 | 4 251 | 201 | 65 951 |
| Niedersachsen | 15 245 | 13 509 | 310 523 | 13 653 | 57 850 | 15 231 | 5 170 | 14 010 | 10 995 | 669 | 456 855 |
| Bremen | 2 513 | 1 841 | 13 625 | 1 252 | 4 094 | 1 569 | 942 | 2 038 | 2 214 | 86 | 30 174 |
| Nordrhein-Westfalen | 15 195 | 8 709 | 58 834 | 3 828 | 678 027 | 35 368 | 30 688 | 33 220 | 30 671 | 2 704 | 897 244 |
| Hessen | 3 597 | 2 839 | 13 574 | 1 044 | 28 165 | 225 776 | 17 674 | 20 080 | 18 433 | 1 184 | 332 366 |
| Rheinland-Pfalz | 1 768 | 1 102 | 4 800 | 711 | 30 170 | 20 489 | 124 694 | 16 746 | 8 738 | 7 040 | 216 258 |
| Baden-Württemberg | 5 163 | 3 915 | 12 101 | 1 184 | 27 452 | 22 763 | 16 101 | 481 228 | 54 647 | 1 804 | 626 358 |
| Bayern | 4 144 | 3 106 | 9 562 | 931 | 24 636 | 13 526 | 8 853 | 56 863 | 527 773 | 911 | 650 305 |
| Saarland | 386 | 186 | 755 | 85 | 3 771 | 2 437 | 8 774 | 3 865 | 1 599 | 41 043 | 62 901 |
| Summe | 203 930 | 52 977 | 453 282 | 26 778 | 876 236 | 344 492 | 215 863 | 637 500 | 663 535 | 55 806 | 3 530 399 |

[1]) Ohne Vertriebene. – [2]) *Bei gleichem Herkunfts- und Zielland handelt es sich um Umzüge innerhalb des Landes mit Ausnahme der Umzüge innerhalb der Gemeinden.*

Quelle: Statistisches Jahrbuch der BRD, Ausgabe 1961, Seite 72; und Ausgabe 1971, Seite 55.

### Anwendung der Verfahren am Beispiel einer Wanderungsmatrix

Um die Arbeitsweise beider Schätzverfahren anhand empirischer Daten zu vergleichen, wurde versucht, die Wanderungen zwischen Bundesländern im Jahre 1969 auf der Basis der Wanderungen im Jahre 1959 ex post zu „prognostizieren". Da die tatsächliche Matrix für 1969 bekannt war, konnte überprüft werden, wie stark die Schätzwerte von den tatsächlichen Werten abwichen.

In der ersten Tabelle sind die tatsächlichen Matrizen angegeben. Diese Tabelle enthält auch die nach dem MODOP-Verfahren geschätzte und die nach dem modifizierten wahrscheinlichkeitstheoretischen Verfahren ermittelte Matrix. Bei dem modifizierten wahrscheinlichkeitstheoretischen Verfahren wurde die einfachste Hypothese über den zeitlichen Verlauf der $k_{ij}$ zugrundegelegt, die es gab, nämlich zeitliche Konstanz: Die Korrekturfaktoren $k_{ij}^{1959}$ wurden auf das Jahr 1969 übertragen. Mit diesen Faktoren wurden unter Zugrundelegen der Randverteilungen der Matrix des Jahres 1969 die Elemente der Matrix des Jahres 1969 ermittelt. Dabei ergab sich erwartungsgemäß, daß die geschätzten Elemente mit den vorgegebenen Zeilen- und Spaltensummen annähernd konsistent waren. Um diese Schätzungen mit der nach dem MODOP-Verfahren gewonnenen vergleichen zu können, wurde in einem zweiten Schritt die vollständige Konsistenz hergestellt. Hierfür wurde die inkonsistente Schätzung als Basismatrix aufgefaßt, aus der eine zweite Matrix mit den Zeilen- und Spaltensummen des Jahres 1969 zu bilden war. Zur Lösung dieses Problems konnte das MODOP-Verfahren angewandt werden, ohne daß die Gefahr bestand, daß die Schätzwerte verfälscht wurden: Da die Konsistenz schon fast erreicht war, fungierte das MODOP-Verfahren in diesem Fall nur noch als Programm zur Aufteilung der geringen Differenzen auf die einzelnen Elemente.

Um die Güte der beiden Schätzverfahren vergleichen zu können, wurde für jedes Verfahren ermittelt, wieviel Prozent der insgesamt gewanderten 3 530 402 Personen den einzelnen Wanderungsströmen richtig

**Matrix der Korrekturfaktoren 1959 und 1969 (relative Abweichungen der tatsächlichen von den hypothetischen Elementen)**

| | Schleswig-Holstein | Hamburg | Niedersachsen | Bremen | Nordrhein-Westfalen | Hessen | Rheinland-Pfalz | Baden-Württemberg | Bayern | Saarland |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Matrix der Korrekturfaktoren $k_{ij}^{1959}$ | | | | | |
| Schleswig-Holstein | − 0,92 | − 0,84 | 0,82 | − 0,47 | 2,66 | 4,85 | 6,03 | 5,80 | 7,33 | 15,01 |
| Hamburg | − 0,85 | – | − 0,46 | − 0,65 | 0,71 | 0,46 | 2,23 | 1,42 | 1,88 | 3,59 |
| Niedersachsen | 0,65 | − 0,50 | − 0,81 | − 0,75 | 0,91 | 1,81 | 4,48 | 4,54 | 6,63 | 8,36 |
| Bremen | − 0,36 | − 0,76 | − 0,72 | − 0,81 | 0,74 | 0,75 | 0,93 | 1,45 | 3,11 | 3,74 |
| Nordrhein-Westfalen | 2,35 | 0,57 | 1,03 | 0,81 | − 0,67 | 1,45 | 0,87 | 3,73 | 4,54 | 3,68 |
| Hessen | 3,79 | 0,63 | 1,98 | 1,23 | 1,69 | − 0,87 | 0,10 | 1,65 | 2,12 | 2,62 |
| Rheinland-Pfalz | 5,63 | 1,96 | 4,94 | 1,35 | 0,77 | 0,01 | − 0,89 | 1,24 | 3,64 | − 0,57 |
| Baden-Württemberg | 5,94 | 1,46 | 5,95 | 3,12 | 4,74 | 1,68 | 1,51 | − 0,77 | 1,19 | 3,50 |
| Bayern | 8,15 | 2,28 | 8,31 | 4,54 | 5,77 | 2,18 | 3,83 | 1,06 | − 0,76 | 9,33 |
| Saarland | 11,76 | 0,26 | 14,38 | 7,10 | 4,80 | 2,48 | − 0,36 | 2,98 | 9,40 | − 0,97 |
| | | | | | Matrix der Korrekturfaktoren $k_{ij}^{1969}$ | | | | | |
| Schleswig-Holstein | − 0,91 | − 0,86 | 0,79 | − 0,32 | 3,18 | 3,73 | 6,08 | 5,43 | 6,40 | 10,29 |
| Hamburg | − 0,88 | – | − 0,43 | − 0,63 | 1,39 | 1,02 | 3,02 | 2,04 | 2,13 | 5,46 |
| Niedersachsen | 0,86 | − 0,49 | − 0,82 | − 0,76 | 1,55 | 2,32 | 5,28 | 5,21 | 6,51 | 9,98 |
| Bremen | − 0,35 | − 0,68 | − 0,76 | − 0,75 | 1,15 | 1,01 | 2,30 | 2,35 | 2,82 | 6,21 |
| Nordrhein-Westfalen | 2,98 | 0,88 | 1,54 | 0,92 | − 0,68 | 1,74 | 0,97 | 3,67 | 4,45 | 3,91 |
| Hessen | 4,64 | 0,75 | 2,76 | 1,30 | 2,09 | − 0,86 | 0,27 | 1,82 | 2,18 | 2,27 |
| Rheinland-Pfalz | 5,77 | 2,13 | 6,21 | 2,69 | 1,05 | 0,19 | − 0,90 | 1,11 | 3,74 | − 0,46 |
| Baden-Württemberg | 6,24 | 1,67 | 6,49 | 2,32 | 4,40 | 1,68 | 1,26 | − 0,77 | 1,44 | 2,26 |
| Bayern | 7,40 | 2,14 | 8,19 | 3,25 | 5,65 | 2,34 | 4,30 | 1,43 | − 0,77 | 6,69 |
| Saarland | 8,68 | 2,96 | 7,92 | 3,44 | 2,70 | 1,24 | − 0,52 | 1,48 | 4,22 | − 0,97 |

zugeordnet werden konnten. Hierfür wurden die Absolutbeträge der Differenzen zwischen den geschätzten und den tatsächlichen Elementen kumuliert. In der auf diese Weise gebildeten Summe war jede falsche Zuordnung zweimal enthalten, denn wenn in einem Land die Zuzüge aus einem Land A zu hoch geschätzt waren, so mußten notwendigerweise die Zuzüge aus einem anderen Land B oder aus mehreren Ländern um das gleiche Ausmaß zu niedrig geschätzt worden sein, weil die Summe der Zuzüge vorgegeben war. Die Gesamtzahl der Personen, die falsch zugeordnet waren, ergab sich daher aus einer Halbierung der kumulierten Differenzen. Beim MODOP-Verfahren waren dies 179 805 oder 2,55 vH aller Personen, beim alternativen Ansatz 182 567 oder 2,59 vH aller Personen. Es ist also festzustellen, daß beide Verfahren recht gute Ergebnisse liefern; sie unterscheiden sich voneinander nur unwesentlich. Aber es ist offenkundig, daß durch eine Trendextrapolation der $k_{ij}$ das Ergebnis des modifizierten wahrscheinlichkeitstheoretischen Verfahrens verbessert werden kann, selbst wenn die Trendfunktionen relativ instabil sein sollten, denn die Extrapolation eines noch so ungesicherten Trends ist in der Regel immer der schematischen Hypothese vorzuziehen, daß sich die $k_{ij}$ im Zeitablauf nicht ändern. Diese Möglichkeit, bei der Extrapolation von Trends die Matrizen und damit den Informationsgehalt von mehreren Jahren zu berücksichtigen, kann auch für das MODOP-Verfahren fruchtbar gemacht werden, denn seine Anwendung setzt nicht voraus, daß bei der Wahl der Ausgangswerte für die Iteration schematisch von Formel (1) ausgegangen wird.

## Summary

This study deals with the problem described as follows: For any given base point in time, a matrix describing certain conditions is known. For a second point in time, only parts of this same matrix are known, namely the sums of the rows and columns. From this information the elements of the complete matrix are to be estimated. The problem can be solved using the RAS procedure or the MODOP procedure. The results from the MODOP procedure depend on the values used to start the iteration process which is to solve the unknown elements of the matrix. Therefore, a means of arriving at the most suitable starting values is offered. This suggested procedure calls for finding the relative deviation between an element of the base matrix and the hypothetical value this same element would have if stochastic independence could be assumed. Then, on the basis of hypotheses on the development of these relative deviations over time, the unknown elements can be estimated from the sums of the rows and columns. The estimation procedure was demonstrated using the example of migration among the states of the Federal Republik of Germany.