

## *LISA (Local Indicators of Spatial Association)*

Prof. Dr. Reinhold Kosfeld, Prof. Dr. Hans-Friedrich Eckey und Matthias Türck, M. A., Kassel

### 1. Einführung

Bei regionalökonomischen Fragestellungen möchte man häufig herausfinden, wie stark sich Regionen gegenseitig beeinflussen, ob also interregionale Einflüsse bei einer Variablen bestehen. Eine solche **räumliche Autokorrelation** tritt aber auch häufig in Regressionsmodellen auf. Das hat zur Folge, dass die Regressionskoeffizienten verzerrt sind oder die Signifikanztests ihre Gültigkeit verlieren (s. Anselin 1988, S. 57 ff.).

Meistens wird bei ökonomischen Fragestellungen eine positive räumliche Autokorrelation vorliegen. Dieser interregionale Einfluss lässt sich inhaltlich auf Spillover-Effekte zurückführen, die in der endogenen Wachstumstheorie und der Theorie der innovativen Netzwerke begründet werden. Spillover-Effekte stellen dabei nicht monetär erfasste Größen von anderen Regionen dar, die die Produktion einer Region erhöhen. Hierbei handelt es sich um Überschwappeffekte beim Wissen oder beim Humankapital.

Eine räumliche Autokorrelation kann mit globalen oder lokalen Maßen geprüft werden. Manche lokalen Maße besitzen die Eigenschaft, dass sie proportional zum Durchschnitt oder zur Summe einer globalen Kennzahl sind. Im Englischen hat sich der Begriff LISA (Local Indicators of Spatial Association) für die lokalen Maße durchgesetzt. Die LISA-Kennzahlen messen für jede einzelne Region die nachbarschaftlichen Beziehungen. Die inzwischen breite Anwendung dieser Maße begründet sich damit, dass räumliche Abhängigkeiten meistens nicht in allen Regionen gleich groß (homogen), sondern lokal unterschiedlich ausgeprägt sind. Man spricht dann von einer **räumlichen Nichtstationarität**. Es gibt also häufig Gruppen oder Cluster von Regionen mit über- oder unterdurchschnittlichen Werten einer Variablen (räumliche Konzentration) und gleich- oder entgegengesetzter Orientierung (positiver und negativer räumlicher Autokorrelation) nebeneinander. Diese kann man mit Hilfe von LISA-Berechnungen identifizieren. Im Folgenden werden die beiden wichtigsten LISA-Maße erläutert und an einem einfachen Beispiel anschaulich berechnet sowie interpretiert.

### 2. Lokaler Moran-Koeffizient

Der lokale Moran-Koeffizient wurde von Anselin (1995, S. 98 ff.) entwickelt. Diese Maßzahl geht von einer Nachbarschaftsmatrix  $\mathbf{W}$  aus.  $\mathbf{W}$  gibt im einfachsten Fall an, ob zwei Regionen  $i$  und  $j$  aneinander angrenzen:

$$(1) \quad w_{ij}^* = \begin{cases} 1, & \text{falls } i \text{ und } j \text{ eine gemeinsame Grenze haben und } i \neq j \text{ ist} \\ 0, & \text{sonst} \end{cases} .$$

Die Interpretation des lokalen Moran-Koeffizienten wird aber verbessert, wenn man die **Nachbarschaftsmatrix**  $\mathbf{W}^*$  so normiert, dass die Zeilensumme eins ergibt. Man teilt hierfür jedes Element durch seine Zeilensumme:

$$(2) \quad w_{ij} = \frac{w_{ij}^*}{\sum_{j=1}^n w_{ij}^*} .$$

Dadurch erhält man die standardisierte Gewichtungsmatrix  $\mathbf{W}$ . Ähnlich wie bei der Berechnung des Korrelationskoeffizienten nach Bravais und Pearson werden zur

Ermittlung des lokalen Moran-Koeffizienten Kreuzprodukte herangezogen. Hier werden aber nicht die Abweichungen zweier unterschiedlicher Variablen X und Y von ihrem Mittelwert verwendet:

$$(3) \quad (x_i - \bar{x}) \cdot (y_i - \bar{y}),$$

sondern einer einzigen Variablen. Statt einer zweiten Variablen Y setzt man das arithmetische Mittel aus den angrenzenden Gebieten ein:

$$(4) \quad \sum_{j=1}^n w_{ij} \cdot (x_j - \bar{x}).$$

Gemessen wird das Produkt der Abweichungen vom Mittelwert in der i-ten Region multipliziert mit dem Mittelwert aller Abweichungen in den Nachbarregionen:

$$(5) \quad (x_i - \bar{x}) \sum_{j=1}^n w_{ij} \cdot (x_j - \bar{x}).$$

Kommen wir zur Erläuterung dieses Kreuzproduktes aus Formel (5), das das Vorzeichen und damit die Interpretation vom lokalen Moran-Koeffizienten bestimmt:

- a) Das Kreuzprodukt in (5) nimmt dann hohe positive Werte an, wenn in der i-ten Region sowie in den angrenzenden Regionen überdurchschnittliche oder im umgekehrten Fall unterdurchschnittliche x-Werte vorhanden sind.
- b) Entsprechend ist das Kreuzprodukt negativ, wenn der x-Wert in der i-ten Region sowie sein arithmetisches Mittel in den Nachbarregionen bei gegenteiligen Vorzeichen vom Mittelwert abweichen.
- c) Ein Kreuzprodukt in der Nähe von null ergibt sich dann, wenn der x-Wert in der i-ten Region oder der Durchschnitt der x-Werte in den angrenzenden Regionen annähernd dem Gesamtmittel entsprechen.

Dieses Kreuzprodukt muss noch normiert werden. Man erreicht dieses, indem man es auf die durchschnittliche Abweichungsquadratsumme von X bezieht. Die Berechnungsformel für den **lokalen Moran-Koeffizienten** der i-ten Region lautet demnach (Fotheringham/Brunsdon/Charlton 2000, S. 102):

$$(6) \quad I_i = \frac{(x_i - \bar{x}) \sum_{j=1}^n w_{ij} \cdot (x_j - \bar{x})}{\sum_{j=1}^n (x_j - \bar{x})^2 / n}.$$

Bei vielen Regionen würde man bei keiner Autokorrelation einen Wert von  $I_i$  nahe bei null erwarten. Bei wenigen Regionen kann sein Erwartungswert, der bei einer standardisierten Gewichtungsmatrix durch

$$(7) \quad E(I_i) = -\frac{\sum_{j=1}^n w_{ij}}{n-1} = -\frac{1}{n-1}$$

gegeben ist, auch merklich von null im negativen Bereich abweichen. Empirische Studien zeigen, dass Spillover-Effekte mit räumlicher Distanz stark abnehmen (vgl. Audretsch/Feldman 2004, Bretschger 1999, S. 252 u. Fritsch 2004, S. 831-832). Damit begründet sich auch die häufige Verwendung der binären Nachbarschaftsmatrix direkt angrenzender Regionen (Nachbarschaftsmatrix erster Ordnung). Selten wird in empirischen Untersuchungen auf eine Nachbarschaftsmatrix zweiter Ordnung

zurückgegriffen, die Effekte zwischen Regionen misst, die durch eine dritte Region voneinander getrennt sind.

Da die Erwartungswerte des lokalen und globalen Moran-Koeffizienten übereinstimmen, kann der globale Moran-Koeffizient als arithmetisches Mittel der lokalen Kennzahlen berechnet werden:

$$(8) \quad I = \frac{1}{n} \sum_{i=1}^n I_i .$$

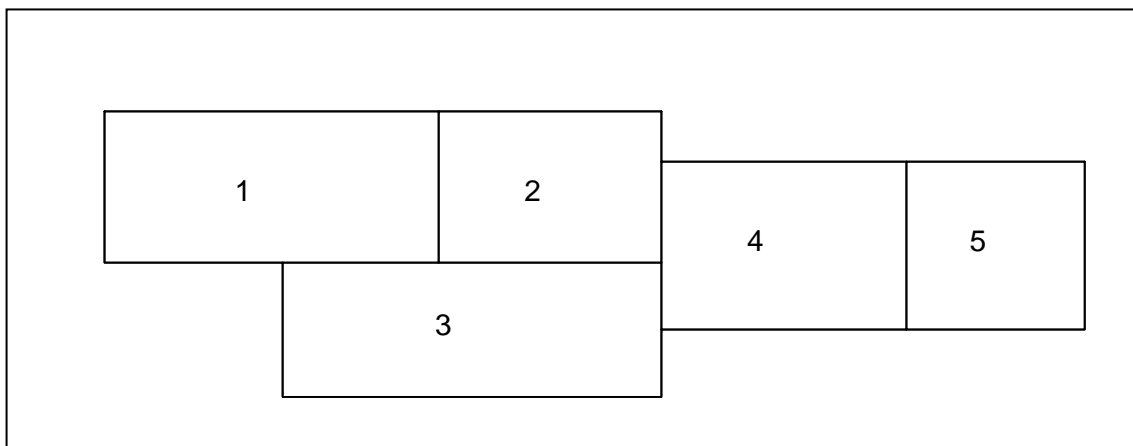
Der globale Moran-Koeffizient  $I$ ,

$$(9) \quad I = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot \sum_{j=1}^n w_{ij} \cdot (x_j - \bar{x})}{\sum_{j=1}^n (x_j - \bar{x})^2} ,$$

liegt bei Verwendung der standardisierten Gewichtungsmatrix fast immer im Intervall zwischen minus und plus eins. Er lässt damit erkennen, ob die lokale Autokorrelation im Mittel über- oder unterdurchschnittlich ausgeprägt ist. Eine genauere Untersuchung erfolgt aber mit den lokalen Kennzahlen. Wenn die lokalen Moran-Koeffizienten grafisch in eine Landkarte eingezeichnet werden, lassen sich räumliche Instationaritäten identifizieren. Hierbei handelt es sich um zusammenhängende Regionen, die im Vergleich zum globalen Moran-Koeffizienten eine stärkere Clusterbildung oder eine entgegengesetzte räumliche Autokorrelation ("Ausreißer") aufweisen.

Verdeutlichen wir uns die Berechnung des lokalen Moran-Koeffizienten an einem Beispiel. Wir gehen dabei von den in Abb. 1 dargestellten fünf Regionen aus.

Abb. 1: Grafische Darstellung der Lage von den Regionen



Für die fünf Regionen muss im ersten Schritt die Nachbarschaftsmatrix  $\mathbf{W}^*$  aufgestellt werden. Die Region 1 grenzt an die Region 2 an, deshalb sind für die Elemente (1,2) und (2,1) eine 1 einzutragen. Entsprechend ist für die übrigen Regionen vorzugehen, wobei auf der Hauptdiagonalen Nullen stehen. Die Nachbarschaftsmatrix für das Fünf-Regionen-Beispiel hat damit folgendes Aussehen:

$$\mathbf{W}^* = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{matrix} & \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \end{matrix} \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix}$$

Anschließend müssen die Zeilensummen ermittelt werden. Die Summe der ersten Zeile beträgt beispielsweise

$$\sum_{j=1}^5 w_{1j}^* = 0 + 1 + 1 + 0 + 0 = 2.$$

Jedes Element der ersten Zeile ist damit durch zwei zu dividieren. Führt man diese Berechnungen für alle Zeilen durch, dann erhält man die standardisierte Gewichtungsmatrix

$$\mathbf{W} = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 & 0 \\ 1/3 & 1/3 & 0 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Neben der standardisierten Gewichtungsmatrix  $\mathbf{W}$  benötigen wir die  $x$ -Werte für die Regionen. Uns interessiert die Frage, wie stark der Preis eines Produktes in einer Region von den entsprechenden Werten in den angrenzenden Gebieten abhängt. Der Preis in Region 1 beträgt 8 Euro, in den Regionen 2 und 3 jeweils 6 Euro etc. Die Preise für alle fünf Regionen sind dem Vektor  $\mathbf{x}$  zu entnehmen:

$$\mathbf{x} = [8 \quad 6 \quad 6 \quad 3 \quad 2]'$$

Das arithmetische Mittel der Preise in den fünf Regionen liegt bei

$$\bar{x} = \frac{1}{5} \cdot \sum_{i=1}^5 x_i = \frac{1}{5} \cdot (8 + 6 + 6 + 3 + 2) = \frac{1}{5} \cdot 25 = 5.$$

Damit lässt sich die Summe der quadrierten Abweichungen vom arithmetischen Mittel [= für Nenner in Formel (6)] bestimmen:

$$\sum_{j=1}^5 (x_j - \bar{x})^2 = (8-5)^2 + (6-5)^2 + (6-5)^2 + (3-5)^2 + (2-5)^2 = 24.$$

Wie lautet dann der lokale Moran-Koeffizient für die erste Region? Das arithmetische Mittel der Abweichungen von den Nachbarregionen ist durch

$$\sum_{j=1}^5 w_{1j} \cdot (x_j - \bar{x}) = 0 \cdot (8-5) + \frac{1}{2} \cdot (6-5) + \frac{1}{2} \cdot (6-5) + 0 \cdot (3-5) + 0 \cdot (2-5) = 1$$

gegeben. Weil nicht-benachbarte Regionen in der ersten Zeile der Nachbarschaftsmatrix durch eine Null gekennzeichnet sind, werden diese Gebiete nicht berücksichtigt. Man erhält damit folgenden lokalen Moran-Koeffizienten:

$$I_1 = \frac{(x_1 - \bar{x}) \sum_{j=1}^5 w_{1j} \cdot (x_j - \bar{x})}{\sum_{j=1}^5 (x_j - \bar{x})^2 / 5} = \frac{(8-5) \cdot 1}{24/5} = \frac{3}{4,8} = 0,6250.$$

Die Region 1 weist also – verglichen mit ihren Nachbarregionen – eine gleichgerichtete Abweichung vom Durchschnittspreis auf. Genauer ist der Preis mit 8 Euro in der Region 1 sowie jeweils 6 Euro in den angrenzenden Regionen 2 und 3 überdurchschnittlich. Wir haben also einen positiven Beitrag zur globalen Autokorrelation, die durch den globalen Moran-Koeffizienten abgebildet wird.

Welchen Wert würde man für den ersten lokalen Moran-Koeffizienten bei fehlender Autokorrelation im Beispiel erwarten? Hier ist Formel (7) anzuwenden:

$$E(I_1) = -\frac{\sum_{j=1}^5 w_{1j}}{5-1} = -\frac{0 + \frac{1}{2} + \frac{1}{2} + 0 + 0}{5-1} = -\frac{1}{4}.$$

Entsprechend zur ersten Region lässt sich der lokale Moran-Koeffizient auch für die zweite Region ermitteln. Wir erhalten diesmal einen Wert von

$$I_2 = \frac{(x_2 - \bar{x}) \sum_{j=1}^5 w_{2j} \cdot (x_j - \bar{x})}{\sum_{j=1}^5 (x_j - \bar{x})^2 / 5} = \frac{(6-5) \cdot (2/3)}{24/5} = \frac{2/3}{4,8} = 0,1389.$$

Entsprechend werden die übrigen lokalen Moran-Koeffizienten bestimmt:

$$I_3 = \frac{2/3}{4,8} = 0,1389, \quad I_4 = \frac{2/3}{4,8} = 0,1389 \quad \text{und} \quad I_5 = \frac{6}{4,8} = 1,2500.$$

Der globale Moran-Koeffizient ergibt sich dann als Durchschnitt der lokalen Maße:

$$I = \frac{1}{5} \cdot \sum_{i=1}^5 I_i = \frac{1}{5} \cdot (0,6250 + 0,1389 + 0,1389 + 0,1389 + 1,2500) = 0,4583.$$

Es liegt also eine deutlich positive Autokorrelation vor. Regionen mit einem überdurchschnittlichen Preis für das Produkt grenzen also tendenziell auch an Gebiete an, bei denen der Preis ebenfalls überdurchschnittlich ausgeprägt ist und vice versa. Die lokalen Moran-Koeffizienten zeigen, dass die Regionen eins und fünf den größten Beitrag zur positiven Autokorrelation liefern. Um die Regionen eins und fünf liegen also Cluster mit ähnlich gerichteten Abweichungen vom Durchschnittspreis.

### 3. Getis-Ord G-Statistiken

Ob lokale Cluster "hot spots" (Konzentration hoher x-Werte) oder "cold spots" (Konzentration niedriger x-Werte) sind, lässt sich anhand der lokalen Moran-Koeffizienten nicht erkennen, da in beiden Fällen eine regional begrenzte räumliche Autokorrelation vorliegt. Hierüber geben die Getis-Ord G-Statistiken Auskunft, die in Getis und Ord (1992) entwickelt worden sind. Diese lokalen Kennzahlen sind bei positiven x-Werten anwendbar, die einen natürlichen Nullpunkt haben (mindestens Verhältnis- oder Ratioskala). Sie gehen von den Distanzen zwischen den Regionen aus, die gewöhnlich durch die Entfernungen oder Fahrtzeiten zwischen den regionalen Zentren gemessen werden. Wir bleiben bei dem Fünf-Regionen-Beispiel (vgl. Abb. 1).

Die Distanz zwischen den Zentren der Regionen eins und zwei liege beispielsweise bei 6 km. Entsprechend werden die übrigen Distanzen ermittelt, die aus Tab. 1 hervorgehen.

Tab. 1: Distanzen zwischen den Regionszentren

Region	1	2	3	4	5
1	0	6	5	11	14
2	6	0	4	5	8
3	5	4	0	7	10
4	11	5	7	0	3
5	14	8	10	3	0

Die Distanzen tragen wir in eine **Distanzmatrix D** ein. Auf der Hauptdiagonale von **D** stehen Nullen, weil die Distanz zwischen dem Zentrum ein und derselben Region bei null liegt:

$$\mathbf{D} = \begin{bmatrix} 0 & 6 & 5 & 11 & 14 \\ 6 & 0 & 4 & 5 & 8 \\ 5 & 4 & 0 & 7 & 10 \\ 11 & 5 & 7 & 0 & 3 \\ 14 & 8 & 10 & 3 & 0 \end{bmatrix}.$$

Im Folgenden wird ein **kritischer Distanzwert d** benötigt, um die Distanzmatrix in eine binäre Matrix (Matrix mit Nullen und Einsen) zu transformieren. Wir wählen hierzu das arithmetische Mittel aller Distanzen ungleich null aus der Distanzmatrix, das bei

$$d = \frac{1}{20} \cdot (6+5+11+14+6+4+5+8+\dots+14+8+10+3) = 7,3$$

liegt. Jetzt ist eine binäre Matrix so zu bestimmen, dass Distanzen, die den kritischen Wert unterschreiten, eins gesetzt werden. Alle übrigen Elemente stellen Nullen dar:

$$(10) \quad w(d)_{ij} = \begin{cases} 1, & \text{falls } d_{ij} < d \\ 0, & \text{sonst} \end{cases}.$$

Im Beispiel beträgt die Distanz zwischen den ersten beiden Regionen  $6 < 7,3$ , weshalb die Elemente (1,0) und (0,1) in der binären Matrix eins sind. Die komplette binäre Matrix hat dann folgendes Aussehen:

$$\mathbf{W}(d = 7,3) = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

In der binären Matrix stehen also Einsen, wenn zwei Regionen relativ nah beieinander liegen, also ihre Distanz die kritische Distanz  $d$  unterschreitet. Bei der Berechnung der einfachen **Getis-Ord-Statistik**  $G_i^*$  wird die Summe der  $x$ -Werte von nah angrenzenden Regionen auf die Summe der  $x$ -Werte aller anderen Regionen bezogen:

$$(11) \quad G_i = \frac{\sum_{j \neq i} w_{ij}(d) \cdot x_j}{\sum_{j \neq i} x_j}.$$

Als Index beim Summenzeichen tritt  $j \neq i$  auf. Damit wird angezeigt, dass über alle Regionen mit Ausnahme der  $i$ -ten Region summiert wird.  $G_i$  gibt also den Anteil der Merkmalssumme der anderen Regionen an, der auf die nicht weit entfernten Regionen zurückgeht. Der Erwartungswert der  $G_i$ -Statistik beträgt

$$(12) \quad E(G_i) = \frac{\sum_{j \neq i} w_{ij}(d)}{n-1}.$$

Übersteigt  $G_i$  diesen Erwartungswert merklich, dann liegt eine positive Autokorrelation mit einer regionalen Konzentration hoher  $x$ -Werte vor. Entsprechend ist eine positive Autokorrelation bei einer lokalen Konzentration niedriger  $x$ -Werte gegeben, wenn  $G_i$  seinen Erwartungswert stark unterschreitet.

Daneben gibt es noch eine Variation dieser Formel, indem man die eigene Region im Zähler und Nenner einbezieht (Haining 2004, S. 253):

$$(13) \quad G_i^* = \frac{\sum_{j=1}^n w_{ij}(d) \cdot x_j}{\sum_{j=1}^n x_j}.$$

$G_i^*$  wäre dann als Anteil der Merkmalssumme aller Regionen zu interpretieren, der auf die eigene Region sowie nicht weit entfernten Gebiete zurückgeht. Der Erwartungswert von  $G_i^*$  beträgt

$$(14) \quad E(G_i^*) = \frac{\sum_{j=1}^n w_{ij}(d)}{n}.$$

Auch bei der Getis-Ord-Statistik gibt es eine entsprechende globale Kennzahl,

$$(15) \quad G = \frac{\sum_{i=1}^n \sum_{j \neq i} w_{ij}(d) \cdot x_i \cdot x_j}{\sum_{i=1}^n \sum_{j \neq i} x_i \cdot x_j},$$

die einen Wertebereich zwischen null und eins aufweist und den Erwartungswert

$$(16) \quad E(G) = \frac{W}{n \cdot (n-1)} \quad \text{mit} \quad W = \sum_{i=1}^n \sum_{j \neq i} w_{ij}(d)$$

hat. Im Gegensatz zum lokalen Moran-Koeffizienten lässt sich  $G$  aber nicht unter Verwendung von  $G_i$  ermitteln. Stattdessen kann man aber auch wieder den globalen Koeffizienten heranziehen und mit Hilfe der lokalen Maße  $G_i$  auf einer Karte "hot spots" und "cold spots" identifizieren.

Die Berechnung soll wiederum am Fünf-Regionen-Beispiel für die Preise illustriert werden. Beginnen wir mit der  $G_i$ -Statistik nach Formel (11). Die  $x$ -Werte (= Preise) lauteten:

$$\mathbf{x} = [8 \ 6 \ 6 \ 3 \ 2]' .$$

Kommen wir zur Berechnung der Zählergröße von (11) für die erste Region. Multipliziert werden die Elemente der ersten Zeile von  $\mathbf{W}(d=7,3)$  mit den  $x$ -Werten. Summiert wird also über die Regionen zwei bis fünf. Die erste Region, für die die Getis-Ord-Statistik  $G_i$  berechnet werden soll, berücksichtigen wir nämlich nicht:

$$\sum_{j \neq 1} w_{1j}(d=7,3) \cdot x_j = 1 \cdot 6 + 1 \cdot 6 + 0 \cdot 3 + 0 \cdot 2 = 12 .$$

Von der Region 1 liegen die Regionen 2 und 3 nicht weit entfernt. Deshalb weist die binäre Matrix  $\mathbf{W}(d=7,3)$  für diese Elemente Einsen auf und nur die Preise dieser beiden Regionen fließt in die Zählergröße ein.

Im Nenner steht die Summe aller  $x$ -Werte, wobei die erste Region wiederum nicht berücksichtigt wird:

$$\sum_{j \neq 1} x_j = 6 + 6 + 3 + 2 = 17 .$$

Die Getis-Ord-Statistik  $G_1$  beträgt damit

$$G_1 = \frac{\sum_{j \neq 1} w_{1j}(d=7,3) \cdot x_j}{\sum_{j \neq 1} x_j} = \frac{12}{17} = 0,7059 .$$

Zirka 71 % der aggregierten Preise aller übrigen Regionen geht auf die Regionen zurück, die nicht weit von Region 1 entfernt liegen. Der Wert von 0,7059 überschreitet erheblich den Wert, den man bei fehlender lokaler Autokorrelation für die erste Region erwarten würde:

$$E(G_1) = \frac{\sum_{j \neq 1} w_{ij}(d=7,3)}{5-1} = \frac{1+1+0+0}{4} = \frac{2}{4} = 0,5 .$$

Um die Region 1 liegt also ein Bündel von Regionen (= Cluster) mit relativ hohen Produktpreisen ("hot spot"). Für die zweite Region erhalten wir beispielsweise eine Getis-Ord-Statistik  $G_i$  von

$$G_2 = \frac{\sum_{j \neq 2} w_{2j}(d=7,3) \cdot x_j}{\sum_{j \neq 2} x_j} = \frac{17}{19} = 0,8947 .$$

Entsprechend kann die zweite Variante der Getis-Ord-Statistik  $G_1^*$  bestimmt werden. Hier fließt jeweils die eigene Region mit ein:



$$G_1^* = \frac{\sum_{j=1}^5 w_{ij}(d=7,3) \cdot x_j}{\sum_{j=1}^5 x_j} = \frac{1 \cdot 8 + 1 \cdot 6 + 1 \cdot 6 + 0 \cdot 3 + 0 \cdot 2}{8 + 6 + 6 + 3 + 2} = \frac{20}{25} = 0,8000.$$

Ein Vergleich mit dem Erwartungswert von

$$E(G_1^*) = \frac{\sum_{j=1}^5 w_{1j}(d=7,3)}{n} = \frac{1+1+1+0+0}{5} = \frac{3}{5} = 0,6$$

zeigt erneut eine Clusterbildung mit hohen (Attributwerten) x-Werten um die Region 1 an. Wie aus der folgenden Tabelle hervorgeht, findet man ähnliche Übereinstimmungen der lokalen Getis-Ord-Maßzahlen vor.

Tab. 2: Getis-Ord-Statistiken

Region i	1	2	3	4	5
$G_i$	$\frac{12}{17} = 0,7059$	$\frac{17}{19} = 0,8947$	$\frac{17}{19} = 0,8947$	$\frac{14}{22} = 0,6364$	$\frac{3}{23} = 0,1304$
$E(G_i)$	$\frac{2}{4} = 0,5$	$\frac{3}{4} = 0,75$	$\frac{3}{4} = 0,75$	$\frac{3}{4} = 0,75$	$\frac{1}{4} = 0,25$
$G_i^*$	$\frac{20}{25} = 0,8000$	$\frac{23}{25} = 0,9200$	$\frac{23}{25} = 0,9200$	$\frac{17}{25} = 0,6800$	$\frac{5}{25} = 0,2000$
$E(G_i^*)$	$\frac{3}{5} = 0,6$	$\frac{4}{5} = 0,8$	$\frac{4}{5} = 0,8$	$\frac{4}{5} = 0,8$	$\frac{2}{5} = 0,4$

Die globale Getis-Ord G-Statistik

$$G = \frac{\sum_{i=1}^5 \sum_{j \neq i} w_{ij}(d=7,3) \cdot x_i \cdot x_j}{\sum_{i=1}^5 \sum_{j \neq i} x_i \cdot x_j}$$

$$= \frac{1 \cdot 8 \cdot 6 + 1 \cdot 8 \cdot 6 + 0 \cdot 8 \cdot 3 + 0 \cdot 8 \cdot 2 + 1 \cdot 6 \cdot 8 + 1 \cdot 6 \cdot 6 + 1 \cdot 6 \cdot 3 + \dots + 1 \cdot 2 \cdot 3}{8 \cdot 6 + 8 \cdot 6 + 8 \cdot 3 + 8 \cdot 2 + 6 \cdot 8 + 6 \cdot 6 + 6 \cdot 3 + 6 \cdot 2 + \dots + 2 \cdot 3}$$

$$= \frac{348}{476} = 0,7311$$

gibt allein global eine Evidenz einer räumlichen Konzentration hoher x-Werte an, da sie ihren Erwartungswert von

$$E(G) = \frac{\sum_{i=1}^5 \sum_{j \neq i} w_{ij}(d=7,3)}{5 \cdot (5-1)} = \frac{12}{20} = 0,6$$

deutlich übersteigt. Die lokalen Maße aus Tab. 2 zeigen an, dass sich diese Cluster ("hot spots") um die ersten drei Regionen konzentrieren.

## 4. Resümee

Eine globale Autokorrelation tritt häufig bei regionalökonomischen Untersuchungen auf. Sie muss aber nicht in allen Regionen nicht nur gleich groß sein, sondern kann auch lokal entgegengesetzt gerichtet sein. Außerdem können sich bei gegebener räumlicher Autokorrelation lokale Cluster hoher ("hot spots") oder niedriger ("cold spots") Werte einer Attributvariablen etablieren. Derartige räumliche Instationaritäten lassen sich mit LISA aufdecken. Vorteilhaft erweist sich dabei, dass mit Hilfe einer kartographischen Darstellung Cluster von Regionen leicht identifiziert werden, die stark wechselseitig voneinander abhängen. Bei weitergehenden Untersuchungen lassen sich für diese Cluster separate Modelle schätzen.

Im Aufsatz wurden die beiden gebräuchlichsten Maße – der lokale Moran-Koeffizient und lokale Getis-Ord-Statistiken – vorgestellt. Die Getis-Ord-Statistiken transformieren eine Distanzmatrix in eine binäre Matrix, womit ein erheblicher Informationsverlust verbunden ist. Treten negative Werte des lokalen Moran-Koeffizienten bei positiver globaler räumlicher Autokorrelation auf, liegen räumliche Instationaritäten in Form von Ausreißern vor ("pockets of instationarity"). Außergewöhnlich hohe lokale Moran-Koeffizienten weisen dagegen auf regionale Cluster hoher ("hot spots") oder niedriger ("cold spots") Werte einer Attributvariablen hin. Um welche der beiden Arten von Clustern es sich dabei handelt, ist aber anhand des lokalen Moran-Koeffizienten nicht erkennbar. Hierzu können die Getis-Ord G-Statistiken eingesetzt werden, anhand derer sich "hot spots" und "cold spots" identifizieren lassen. Ein "hot spot" zeigt sich anhand eines stark überdurchschnittlichen  $G_i$  - oder  $G_i^*$  -Wertes, während ein "cold spot" durch einen deutlich unterdurchschnittlichen  $G_i$  - oder  $G_i^*$  -Wert gekennzeichnet ist.

Für die lokalen Kennzahlen gibt es auch Signifikanztests, die in gängige Software für GIS (Geoinformationssysteme) integriert sind. Die Berechnung dieser Größen ist teilweise recht komplex. Insbesondere muss man berücksichtigen, dass mehrere Signifikanztests gleichzeitig (simultan) durchgeführt werden. Man möchte beispielsweise wissen, inwiefern eine globale Autokorrelation, die signifikant ist, auf einzelne signifikante lokale Autokorrelationen zurückgeführt werden kann. Soll dann das Signifikanzniveau von beispielsweise 5 % für alle lokalen Signifikanztests gleichzeitig gelten, muss eine **Alpha-Fehler-Adjustierung** durchgeführt werden.

## Literatur

- Anselin, L.*, Spatial Econometrics. Methods and Models, Dordrecht, Boston, London 1988.
- Anselin, L.*, Local Indicators of Spatial Association-LISA, Geographical Analysis, Vol. 27(1995), S. 93-115.
- Audretsch, D. B., M. P. Feldman*, Knowledge Spillovers and the Geography of Innovation. In: Henderson J. V., Thisse J.-F. (Hrsg.), Handbook of Regional and Urban Economics, Volume 4, Amsterdam u. a. 2004, S. 2713-2742.
- Bretschger, L.*, Knowledge Diffusion and the Development of Regions, Annals of Regional Science, Vol. 33(1999), S. 251-268.
- Fotheringham, A. S., C. Brunson, M. E. Charlton*, Quantitative Geography. Perspectives on Spatial Data Analysis, London, New Delhi 2000.
- Fritsch, M.*, Cooperation and the Efficiency of Regional R&D Activities, Cambridge Journal of Economics, Vol. 28(2004), S. 829-846.

*Getis, A., J. K. Ord*, The Analysis of Spatial Association by Use of Distance Statistics, Geographical Analysis, Vol. 24(1992), S. 189-206.

*Getis, A., J. K. Ord*, Local Spatial Statistics. An Overview, in: P. Longley, M. Batty (Hrsg.), Spatial Analysis. Modelling in a GIS Environment, New York 1996, S. 261-278.

*Haining, R.*, Spatial Data Analysis. Theory and Practice, Cambridge 2004.