

einer Ausnahme: bestimmungsgemäß erfolgen die Gartenschauinvestitionen in der Stadtmitte im Zusammenhang mit der Stadtsanierung, also in einem Bereich, der bisher bei den Grüninvestitionen zu kurz gekommen ist. Die besondere Anstrengung im Zusammenhang mit der Gartenschau liegt daher in der Entscheidung für Dauergrünanlagen anstelle der bisher geplanten Bauten sowie in dem Bündelungseffekt für andere Investitionen, die sonst vielleicht etwas später vorgenommen worden wären. Nutznießer sind die Bewohner der Innenstadt, denen langjährige Baustellen erspart bleiben und die endlich die ihnen zustehende Gleichbehandlung erfahren.

Damit findet stadtpolitisch ein wichtiger Ausgleich zwischen eingemeindeten Vororten und Kernstadt seinen Abschluß; während der vergangenen zehn Jahre waren Eingemeindungsverpflichtungen zu erfüllen, flossen die Investitionsmittel - auch im Grünbereich - in die Vororte. Nunmehr kann die Kommunalpolitik, von einem gleichmäßigen Niveau der Infrastrukturversorgung ausgehend, mit einer neuen Investitionsrunde beginnen.

Räumliche Querschnittsregressionen mit  
potentialisierten Variablen

von

Johannes Bröcker

Institut für Regionalforschung der  
Christian-Albrechts-Universität Kiel  
Olshausenstraße 40, 2300 Kiel

<u>Inhalt:</u>	<u>Seite</u>
1. Problemstellung	56
2. Modellspezifikation	58
3. Schätzverfahren	64
3.1 Almon-Verfahren	64
3.2 Verfahren mit Regionsringen	69
3.3 Nichtlineare Verfahren	77
3.3.1 Kleinste-Quadrate-Prinzip	79
3.3.2 Vertrauensbereiche und Tests	81
3.3.3 Minimierung der Residuenquadrate	89
4. Schlußfolgerungen	97
Literaturverzeichnis	98

-----  
Das Papier behandelt methodische Aspekte empirischer Arbeiten, die im Rahmen des Sonderforschungsbereichs 17 "Skandinavien- und Ostseeraumforschung" im Teilprojekt C 13 durchgeführt und durch die Deutsche Forschungsgemeinschaft finanziell gefördert wurden.

## 1. Problemstellung

Zu den Standardinstrumenten der empirischen Regionalforschung gehört das Verfahren der linearen Regression, mit dem sich Hypothesen über den Einfluß exogener Variablen (der Regressoren) auf eine oder mehrere endogene Variablen (Regressanden) testen sowie Stärke und Richtung des Einflusses bestimmen lassen. Während in der ökonomischen Analyse makroökonomischer Problemstellungen ganz überwiegend mit Zeitreihen gearbeitet wird, steht in der Regionalökonomie häufig die Untersuchung räumlicher Querschnitte im Vordergrund. Daneben gibt es - abgesehen von reinen Längsschnittanalysen, die natürlich auch auf regionaler Ebene durchgeführt werden - auch kombinierte Quer- und Längsschnittuntersuchungen. Wir wollen uns hier jedoch auf die Diskussion räumlicher Querschnittsregressionen beschränken.

Die Dominanz makroökonomischer Fragestellungen hat in der Ökonometrie und Statistik dazu geführt, daß die Zeitreihenanalyse hier zum Hauptforschungsfeld geworden ist und Probleme wie serielle Fehlerkorrelation, zeitverzögerte Abhängigkeit endogener Variablen von sich selbst, zeitverzögerte Wirkung exogener Variablen etc. extensiv bearbeitet worden sind. Dies hat eine Vielzahl spezieller Modelle und Schätztechniken sowie leicht zugängliches Software (z.B. TSP) speziell für Zeitreihen hervorgebracht.

Die in räumlichen Querschnittsregressionen auftretenden Probleme sind vergleichsweise unterbelichtet. Am weitesten gediehen scheint mittlerweile die Behandlung des mit der seriellen Fehlerkorrelation korrespondierenden Problems räumlicher Fehlerkorrelation. Ausgehend von dem bekannten Beitrag von Cliff und Ord (1973) zu diesem Thema ist inzwischen eine weitverzweigte Literatur herangewachsen, auf die ich hier nicht weiter eingehen will, u.a. weil mir das Problem der Fehlerkorrelation - anders als in der Zeitreihenanalyse - in räumlichen Querschnitten weniger bedeutend erscheint. Kaum bearbeitet sind Regressionsansätze aus dem Bereich räumlicher Querschnittsanalysen, in denen unterstellt ist, daß der Wert der abhängigen Variable  $y_r$  in der Region  $r$  nicht allein

abhängt von erklärenden Variablen, die Eigenschaften eben dieser Region  $r$  messen, sondern auch von solchen, die sich auf Nachbarregionen oder auch weiter entfernt liegende Regionen beziehen. Häufig ist ein mit zunehmender Distanz abnehmender, insgesamt aber nicht zu vernachlässigender Einfluß umliegender Regionen zu unterstellen.

Die hier angedeutete Struktur von räumlichen Querschnittsmodellen ist in vieler Hinsicht verwandt mit der von Zeitreihenmodellen, in denen zeitverzögerte erklärende Variablen auftreten. In Analogie zum time-lag kann daher auch vom space-lag gesprochen werden. Leider können die gut ausgearbeiteten Verfahren, mit denen sich Zeitreihenmodelle mit time-lags schätzen lassen, meist nicht auf die Schätzung von Querschnittsregressionen mit space-lags übertragen werden, weil trotz der Ähnlichkeit beider Probleme Unterschiede bleiben, die sich für die Anwendung der üblichen Verfahren als essentiell erweisen.

Die Zuordnung der Beobachtungen zu verschiedenen Zeitpunkten weist einige Wesenszüge auf, die für die Zuordnung zu Raumpunkten im allgemeinen nicht gelten: Zeitpunkte sind seriell geordnet, und die Wirkungen gehen nur in eine Richtung. Es gibt keine Auswirkung der Zukunft auf die Vergangenheit. Daraus ergibt sich weiterhin, daß für aufeinanderfolgende Zeitpunkte  $t_1$ ,  $t_2$  und  $t_3$  der zeitliche Abstand von  $t_1$  nach  $t_3$  gleich der Summe des Abstandes von  $t_1$  nach  $t_2$  und von  $t_2$  nach  $t_3$  ist. Bisweilen lassen sich auch Raumpunkte seriell ordnen (Tankstellen an einer Autobahn), doch wird man praktisch nie nur eine Wirkungsrichtung unterstellen können. Im allgemeinen kann man Raumpunkte nicht "auf die Schnur ziehen", und die Wirkungen weisen in alle Richtungen. Für den räumlichen Abstand zwischen Punkten  $r_1$  und  $r_3$  gilt im allgemeinen die Dreiecksungleichung, d.h. daß der Abstand zwischen  $r_1$  und  $r_3$  höchstens so groß ist wie der zwischen  $r_1$  und  $r_2$  zuzüglich des Abstandes zwischen  $r_2$  und  $r_3$ . Schließlich unterscheidet sich die Zuordnung zu Zeitpunkten von der zu Raumpunkten auch dadurch, daß regelmäßige Zeitabstände (Monats-, Quartals-, Jahresdaten etc.) die Regel sind, während regelmäßige Regionsabstände (z.B. Beobachtungen auf Rasterpunkten) die

Ausnahme bilden. Bei unregelmäßigen Zeitabständen wäre Zeitreihenanalyse zwar nicht unmöglich, aber doch erheblich schwieriger.

Einige aus der Behandlung von Zeitreihen bekannte einfache Verfahren können auf Querschnittsmodelle mit space-lags übertragen werden, weil sie die speziellen Wesenszüge der zeitlichen Ordnung nicht oder nur teilweise ausnutzen. Diese werden in den ersten beiden Unterabschnitten von Abschnitt 3 erläutert. Die ausgefeilteren Verfahren zur Schätzung von Regressionsmodellen mit Lag-Verteilungen verwenden die autoregressiv transformierte Form dieser Modell, die nur aufgrund der dargestellten speziellen Struktur der zeitlichen Ordnung möglich wird (vgl. z.B. Johnston, 1963, Kapitel 10). Wo die einfachen Verfahren in der Querschnittsanalyse versagen, sind daher spezielle Schätzmethoden zu entwickeln, die Gegenstand des dritten Unterabschnittes von Abschnitt 3 sind.

## 2. Modellspezifikation

Betrachten wird als Beispiel eine lineare Querschnittsregression mit deren Hilfe das Wachstum (der Beschäftigten, des Umsatzes, des BIP oder dergleichen) in der Region  $r$  ( $y_r$ ) durch eine Reihe regionaler Determinanten ( $x_{r1}, \dots, x_{rp}$ ) statistisch erklärt werden soll:

$$y_r = \alpha_1 x_{r1} + \dots + \alpha_i x_{ri} + \dots + \alpha_p x_{rp} + \epsilon_r = x_{r.} \alpha + \epsilon_r \quad (1)$$

$\alpha$  ist ein  $(p \times 1)$ -Spaltenvektor, dessen Komponenten die Regressionskoeffizienten  $\alpha_1$  bis  $\alpha_p$  sind.  $x_{r.} = (x_{r1}, \dots, x_{rp})$  ist der  $(1 \times p)$ -Zeilenvektor der regionalen Wachstumsdeterminanten  $x_{r1}$  bis  $x_{rp}$ .<sup>1</sup>  $\epsilon_r$  ist eine stochastische Störvariable, über deren Eigenschaften gewisse Annahmen getroffen werden müssen, um eine Schätzung zu ermöglichen.

<sup>1</sup> Zur Vereinfachung der Schreibweise wird in der Regression der Achsenabschnitt nicht explizit aufgeführt. In einem inhomogenen Regressionsansatz ist für ein  $i$  zu setzen:  $x_{ri} = 1$  für alle  $r$ .  $\alpha_i$  ist dann der Achsenabschnitt.

Die Regressoren 1 bis  $p$  quantifizieren verschiedene Wachstumsdeterminanten wie

- Zugangsmöglichkeiten zu Output- und Inputmärkten,
- Agglomerationsvor- und -nachteile,
- Möglichkeiten zur Realisierung von Skalenerträgen,
- Verfügbarkeit qualifizierter Arbeitskräfte,
- Infrastrukturausstattung,
- Regionalförderung, etc.

Wachstumsansätze dieser Art wurden u.a. von Koll (1979), Reimers (1981) und Bröcker, Peschel und Reimers (1983) geschätzt. Einen ausführlichen Survey der weiteren einschlägigen Literatur bringt Reimers (1981). Die Betrachtung eines Wachstumsansatzes soll hier nur der Illustration dienen. Die zu erläuternden statistischen Probleme stellen sich in gleicher Weise, wenn man ganz andere Fragen, z.B. den Einfluß regionaler Lebensbedingungen auf Krankenstand oder Lebenserwartung, die Determinanten der Partizipation von Frauen am Erwerbsleben oder was auch immer im Regionsquerschnitt untersuchen will.

In Gleichung (1) tauchen nur Regressoren auf, die denselben Regionsindex  $r$  wie die abhängige Variable tragen. Mit Sicherheit hängt jedoch das Wachstum in der Region  $r$  nicht nur von Bedingungen in dieser Region selbst ab, sondern auch von der Situation in der näheren und weiteren Umgebung. Nur hinsichtlich weniger der oben angeführten Wachstumsdeterminanten dürfen wir die Region als eine isolierte Einheit betrachten. Die Möglichkeiten zur Realisierung von Skalenerträgen etwa werden sinnvollerweise durch Indikatoren gemessen, die sich allein aus den Größen der Betriebe errechnen, die ihren Standort in der betrachteten Region haben (oder aus den Größen der Unternehmen, zu denen diese Betriebe gehören). Betrachtet man aber z.B. die Verfügbarkeit von qualifizierten Arbeitskräften, so wird man die Verfügbarkeit nicht nur in der eigenen Region, sondern zumindest im Pendlereinzugsbereich aller regionalen Standorte in Betracht ziehen müssen. Dieser deckt sich

sich umso weniger mit der Region selbst, je kleiner sie abgegrenzt ist und je weiter die Regionsabgrenzung vom Idealbild einer funktional abgegrenzten Arbeitsmarktregion entfernt ist.

Betrachten wir schließlich Marktzugangsmöglichkeiten oder die Betroffenheit der regionalen Produzenten von Agglomerationsvor- oder -nachteilen, so handelt es sich hier wesentlich um interregionale Phänomene. Hinsichtlich der Outputmärkte etwa sind zwar auch die Absatzmöglichkeiten in der eigenen Region von Bedeutung, jedoch nicht allein, und bei hinreichend feiner Regionsgliederung nicht einmal überwiegend. Wichtig sind hier die Möglichkeiten, von der Region r aus die Märkte im ganzen Land oder sogar in der ganzen Welt zu bedienen. Ähnliches gilt für Agglomerationseffekte: Vorteile aufgrund der städtischen Konzentration bestimmter Dienste, Informationen etc. kommen nicht nur Firmen in einer Stadt selbst, sondern auch solchen in ihrer Umgebung zugute.

Erklärende Variablen mit space-lags oder "Fernwirkung" können im Regressionsansatz (1) auf zweierlei Weise berücksichtigt werden. Entweder nimmt man weitere Regressoren, die andere Regionsindices tragen, zusätzlich in die Gleichung auf, oder man faßt einen oder mehrere Regressoren  $x_{ri}$  selbst wieder als Funktionen exogener Variablen auf, die ihrerseits nicht nur Eigenarten der Region selbst messen, sondern auch die anderer Regionen. Wir wählen die zuletzt genannte Darstellung, die sich als flexibler erweist und die zuerst genannte als Sonderfall enthält.

Für  $x_{ri}$  setzen wir also in die Regression ein

$$x_{ri} = \xi_{ri}(\beta_i, z_i) \quad i=1, \dots, q \leq p. \quad (2)$$

$x_{ri}$  wird jetzt zu einer nicht beobachtbaren, latenten Variable, die nach einer bestimmten, aufgrund theoretischer Vorstellungen postulierten Funktion  $\xi_{ri}$  von beobachtbaren exogenen Variablen abhängt, die wir im Vektor  $z_i$  zusammenfassen. Unsere Unkenntnis über die genaue Form des Funktionalzusammenhangs schlägt sich darin nieder, daß  $x_{ri}$  außer von  $z_i$  auch von einem unbekanntem Parameter(-vektor)  $\beta_i$  abhängt, der zusammen mit den  $\alpha$ -Parametern

zu schätzen ist. Wir wollen ohne Beschränkung der Allgemeinheit annehmen, daß die Regressoren für  $i = 1, \dots, q \leq p$  die Form (2) besitzen, während die weiteren Regressoren mit  $i = q+1, \dots, p$  bekannte Größen sind.

Die Funktionen  $\xi$  sind derart zu spezifizieren, daß  $\beta_i$  etwas darüber aussagt, wie die Wirkung einer Variable  $z_{is}$ , die die Ausprägung der Wachstumsdeterminante  $i$  in der Region  $s$  mißt, auf das Wachstum in  $r$  durch die Entfernung zwischen  $r$  und  $s$  beeinflusst wird.  $\beta_i$  bezeichnen wir als Distanzparameter.

Um die Spezifikation dieser Funktionen zu konkretisieren, ist es hilfreich, zwischen zwei Typen erklärender Variablen  $z_i$  zu unterscheiden, die wir als Absolutvariablen auf der einen Seite und Relationsvariablen auf der anderen bezeichnen wollen. Wie der Name schon sagt, sollen Absolutvariablen solche sein, die durch ihre absolute Höhe, nicht durch ihre Relation zu anderen Regionalvariablen das Wachstum beeinflussen. Die Größe des Marktes, die Zahl qualifizierter Arbeitskräfte etc. sind derartige Variablen. Der Wert solcher Variablen steigt tendenziell mit der Regionsgröße. Bei diesem Variablentyp ist eine kumulative Wirkung der von verschiedenen Regionen ausgehenden Wachstumseinflüsse anzunehmen. Betrachten wird - anders gesagt - den Effekt von Variablen  $z_{is}$ ,  $s \in U_r$ , auf das Wachstum in  $r$ , wobei  $z_{is}$  eine wachstumsbeeinflussende Eigenart der Region  $s$  mißt und  $U_r$  eine irgendwie abgegrenzte Umgebung von  $r$  bezeichnet.<sup>1</sup> Dann wird bei Absolutvariablen unterstellt, daß die Wirkungen der Variablen  $z_{is}$  für die verschiedenen  $s$  sich addieren, wobei die Stärke des Einflusses von  $z_{is}$  auf das Wachstum in  $r$  von der Entfernung zwischen  $r$  und  $s$  abhängt.

Die Spezifikation der  $\xi$ -Funktion als kumulatives Potential führt auf eine geeignete Form zur Repräsentation von Absolutvariablen

<sup>1</sup>  $U_r$  kann auch die Menge aller Region des Untersuchungsraumes bezeichnen, die für alle  $r$  identisch ist. Die Abgrenzung der regionen  $s \in U_r$ , auf die sich  $z_{is}$  bezieht, braucht mit der, auf die sich  $y_r$  bezieht, im übrigen nicht identisch zu sein.

in der Regressionsgleichung. Als kumulatives Potential definieren wir

$$\xi_{ri}(\beta_i, z_i) = \sum_{s \in U_r} a(\beta_i, c_{rs}) \cdot z_{is} \quad (3)$$

Die durch  $\beta_i$  parametrisierte, von der Entfernung  $c_{rs}$  zwischen  $r$  und  $s$  abhängige Funktion  $a$  bezeichnen wir als Zugangsfunktion.  $\xi_{ri}$  hängt natürlich außer von  $\beta_i$  und  $z_i$  auch von den  $c_{rs}$ ,  $s \in U_r$ , ab, was wir aber nicht besonders hervorheben. Im folgenden schreiben wir statt  $\xi_{ri}(\beta_i, z_i)$  auch einfach  $\xi_{ri}(\beta_i)$ , wenn die Abhängigkeit von  $z_i$  nicht besonders hervorgehoben werden soll.

Setzt man (3) für  $x_{ri}$  in die Regression (1) ein, erhält man

$$Y_r = \alpha_1 x_{r1} + \dots + \sum_{s \in U_r} \tilde{\alpha}_{irs} z_{is} + \dots + \alpha_p x_{rp} + \varepsilon_r \quad (4)$$

$$\text{mit } \tilde{\alpha}_{irs} = \alpha_i a(\beta_i, c_{rs}), \quad (5)$$

also eine neue Regression mit den beobachtbaren Regressoren  $z_{is}$  und den Regressionskoeffizienten  $\tilde{\alpha}_{irs}$ , die über (5) vom Parameter  $\beta_i$  und den Distanzen abhängen. Die Zugangsfunktion  $a$  gibt also an, wie der Einfluß, der durch die Wachstumsdeterminante  $i$  in der Region  $s$  auf das Wachstum in  $r$  ausgeübt wird, von der Entfernung zwischen  $r$  und  $s$  abhängt.

Kommen wir zum zweiten Variablen-Typ, den wir als Relationsvariable bezeichnen. Darunter verstehen wir Variablen, die in Form von regionalen Relationen oder Durchschnitten gemessen werden, wie etwa Lohnsätze, Bodenpreise, Pro-Kopf-Einkommen, Steuersätze etc. Bei solchen Variablen ist es offensichtlich sinnlos, die von verschiedenen Regionen ausgehenden Wirkungen zu kumulieren. Dennoch sind auch hier space-lags zu berücksichtigen. Die wachstumshemmende Wirkung hoher Lohnsätze etwa schwächt sich ab, wenn in umliegenden Regionen im Pendlereinzugsbereich niedrigere Löhne als in  $r$  selbst vorgefunden werden. Auch hier wird die Bedeutung von Bedingungen in anderen Regionen davon abhängen, wie weit diese entfernt sind. Die Wirkung von Relationsvariablen mit space-lags läßt sich durch die Spezifikation der  $\xi$ -Funktion als

Durchschnittspotential modellieren, das wie folgt definiert wird

$$\xi_{ri}(\beta_i, z_i) = \sum_{s \in U_r} w_r(\beta_i, c_{rs}) \cdot z_{is} \quad (6)$$

$$\text{mit } w_r(\beta_i, c_{rs}) \geq 0 \quad \text{und} \quad \sum_{s \in U_r} w_r(\beta_i, c_{rs}) = 1. \quad (7)$$

$w_r$  bezeichnen wir als Gewichtungsfunktion. Der formale Unterschied zwischen dem kumulativen Potential (3) und dem Durchschnittspotential (6) besteht darin, daß  $w_r$  im Gegensatz zu  $a$  den Beschränkungen (7) unterliegt. Dieser Unterschied ist wesentlich für die Verwendbarkeit der einfachen Verfahren, die unter 3.1 und 3.2 darzustellen sind. Er impliziert, daß die Gewichtungsfunktion  $w_r$  sich im Gegensatz zur Zugangsfunktion  $a$  von Region zu Region unterscheiden muß. Anderenfalls könnte die zweite der Restriktionen (7) im allgemeinen nicht erfüllt werden.

Mit der Spezifikation der Zugangs- und Gewichtungsfunktion befassen wir uns erst im nächsten Abschnitt im Zusammenhang mit den konkreten Schätzproblemen. Hier ist noch zu betonen, daß Regressionen mit latenten Regressoren, die, wie in (2) angedeutet, von unbekanntem Distanzparametern und bekannten Variablen abhängen, natürlich außer den genannten eine Vielzahl anderer Formen zulassen, in denen regionsübergreifende Wirkungsketten, die in a priori unbekannter Weise durch die Distanz beeinflußt werden, modelliert werden können.

Hinzuweisen ist hier zum einen auf Konsumentensurplus-Indikatoren, die aus partiellen Gleichgewichtsmodellen räumlicher Interaktionen hergeleitet werden. Als Accessibility- oder User-Benefit-Maße spielen sie in der verkehrstheoretischen Literatur (Williams, 1977; Ben-Akiva und Lerman, 1979) eine wichtige Rolle. Dieselben Maße können auch im Rahmen interregionaler Handelsmodelle formuliert werden und führen hier zur Ersetzung herkömmlicher Potentialmaße durch theoretisch besser begründete Marktzugangs-Indikatoren (Bröcker, 1980; 1984). Diese Indikatoren hängen in verwickelter Form vom Angebot und der Nachfrage in sämtlichen Regionen sowie den interregionalen Distanzen zwischen allen Regionen des Untersuchungsraumes ab. Außerdem hängen sie

von einem Distanzparameter ab, der den Einfluß von Entfernungsdifferenzen auf die Anbieter-Wahl charakterisiert, und können daher in der Form (2) geschrieben werden. Nach den im Abschnitt 3.3 darzustellenden Verfahren können daher zusammen mit den Regressionsparametern auch die Distanzparameter der Interaktionsmodelle geschätzt werden, die den als Regressoren verwendeten Accessibility-Maßen zugrunde liegen.

Hinzuweisen ist ferner auf eine Reihe meist mehrparametrischer Maße, mit denen sich die Stellung einer Region im räumlichen Gefüge des Siedlungssystems charakterisieren läßt. Als Beispiel seien hier die verschiedenen Agglomerations- und Lagemaße erwähnt, die Koll (1977), ausgehend von theoretischen Überlegungen Richardsons (1973) und v. Böventers (1975), in seinem Wachstumsansatz verwendet. Koll arbeitet mit schwer begründbaren, willkürlichen Parametervorgaben. Mit den nichtlinearen Methoden des Abschnitts 3.3 ließen sich auch diese Parameter an Hand von Beobachtungen schätzen.

### 3. Schätzverfahren

#### 3.1 Almon-Verfahren

In der Zeitreihenanalyse wird häufig unterstellt, die Lag-Verteilung, das zeitliche Pendant zur Zugangs- bzw. Gewichtungsfunktion, lasse sich durch einen Polynom vorgegebenen Grades darstellen (sogenannter Almon-Lag; vgl. z.B. Johnston, 1963, S. 294). Bekanntlich läßt sich jede stetige Funktion auf einem Intervall durch einen Polynom hinreichend hohen Grade beliebig gut approximieren (Approximationssatz von Weierstrass, vgl. z.B. Reinhardt und Soeder, 1977).

Das Verfahren läßt sich auf die Schätzung eines Querschnittsansatzes mit kumulativen Potentialen übertragen und führt auf ein

einfaches lineares Schätzproblem. Mit Durchschnittspotentialen gibt es allerdings Schwierigkeiten.

Gehen wir also zuerst davon aus, daß nur kumulative Potentiale auftreten und nehmen an, die Zugangsfunktion lasse sich in einem Intervall  $[0, c_{\max}]$  durch ein Polynom vom Grade  $v$  darstellen:

$$a(\beta_i, c_{rs}) = \sum_{u=0}^v \beta_{iu} c_{rs}^u \quad (8)$$

Für  $c_{rs} > c_{\max}$  wird  $a(\beta_i, c_{rs}) = 0$  gesetzt. Dadurch ist die Umgebung  $U_r$  implizit als Menge der Regionen definiert, deren Entfernung zur Region  $r$  kleiner oder gleich  $c_{\max}$  ist.  $c_{\max}$  und  $v$  sind a priori vorzugeben.

Setzt man (8) in (3) und (3) für  $x_{ri}$  in (1) ein, geht (1) über in

$$y_r = \alpha_1 x_{r1} + \dots + \sum_{u=0}^v \gamma_{iu} \left\{ \sum_{s \in U_r} c_{rs}^u z_{is} \right\} + \dots + \alpha_p x_{rp} + \epsilon_r \quad (9)$$

Der Summand in der ursprünglichen Regression wird also ersetzt durch die Summe von  $v+1$  neuen Regressoren (die Ausdrücke in  $\{\}$ ), jeweils versehen mit den Regressionskoeffizienten  $\gamma_{iu}$  mit  $\gamma_{iu} = \alpha_i \beta_{iu}$ .

Die neuen Regressoren lassen sich ohne Kenntnis der Parameter aus den Daten berechnen, und die neuen Regressionskoeffizienten nach den üblichen Verfahren<sup>2</sup> schätzen. Das macht dieses Verfahren, das wir als Almon-Verfahren bezeichnen wollen, natürlich attraktiv.

Das Problem dieses Ansatzes ist, daß die Spezifikation (8) nicht erlaubt, dem Verlauf der Zugangsfunktionen Restriktionen aufzuerlegen, die a priori plausibel erscheinen, wie beispielsweise die

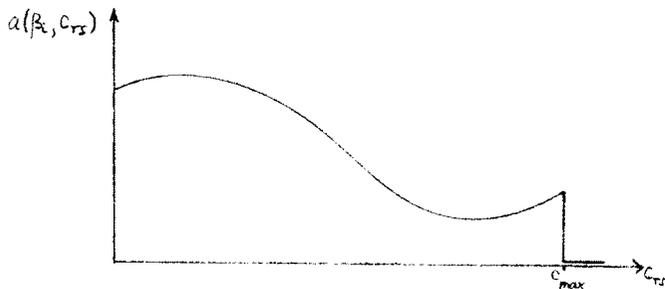
1 Man beachte, daß  $u$  in  $c_{rs}^u$  einen ganzzahligen Exponenten, nicht einen hochgestellten Index bezeichnet.

2 Mit den üblichen Verfahren sind hier gewöhnliche Kleinste-Quadrate (OLS) oder ein- und zweistufige Aitken-Schätzer (GLS) gemeint, je nach dem, welche Annahmen über Residuen-Varianzen und -Kovarianzen getroffen werden.

Restriktion, daß die Funktion monoton sinken und für  $c_{rs} \rightarrow \infty$  gegen Null streben soll. Tatsächlich wird die Zugangsfunktion (8) immer gegen  $-\infty$  oder  $+\infty$  streben wenn  $c_{rs}$  über alle Grenzen wächst. Daher läßt sich nicht vermeiden, die Zugangsfunktion von vornherein auf Bereiche unterhalb einer vorgegebenen Maximaldistanz  $c_{max}$  zu beschränken. Das bedeutet, daß die räumliche Reichweite der im Potential erfaßten Variable, deren Schätzung ja eigentlich der Sinn des ganzen Verfahrens ist, zunächst hinsichtlich ihres Maximums a priori vorgegeben wird. Lediglich der Funktionsverlauf innerhalb des Maximalradius wird geschätzt.

Die Spezifikation (8) läßt Funktionsverläufe zu, die ökonomisch wenig plausibel sind. Bei der Schätzung regionalökonomischer Modelle wird es häufig zu unplausiblen geschätzten Funktionsverläufen kommen, weil eine sehr enge Anpassung der Schätzungen an die Beobachtungen hier nicht zu erwarten ist. Die multiple Bestimmtheit in Querschnittsregressionen, mit denen regionale Wachstumsunterschiede erklärt werden sollen, liegt typischerweise in der Größenordnung von 0,5. Bei einem Zugangspolynom dritten Grades wäre z.B. der in Abbildung 1 gezeigte Verlauf kein unwahrscheinliches Schätzergebnis. Die Funktion zeigt ein Maximum und ein Minimum,

Abbildung 1: Zugangspolynom dritten Grades



fallende und steigende Bereiche sowie eine abrupte Stufe an der Grenze  $c_{max}$ , bei deren Überschreitung die Zugangsfunktion annahmegemäß den Wert Null annimmt.

Der unerwünschte Sprung an der Stelle  $c_{max}$  läßt sich vermeiden, wenn man die Parameter unter der Restriktion

$$a(\beta_i, c_{max}) = \sum_{u=0}^v \beta_{iu} c_{max}^u = 0 \quad (10)$$

schätzt. Dies bringt keine schätztechnischen Komplikationen mit sich, da die Restriktion (10) eine lineare Gleichung ist. OLS- und GLS-Schätzer bleiben linear und behalten ihre wünschenswerten Eigenschaften, wenn die Parameter Restriktionen in Form linearer Gleichungen unterworfen werden (Schönfeld, 1971, S. 125-127). Für die Praxis ist es am einfachsten, die Restriktion (10) nach einem Parameter aufzulösen und in die Regression einzusetzen, so daß man einen unbeschränkten linearen Ansatz erhält. Man multipliziert (10) mit  $\alpha_i$  und löst nach  $\gamma_{io}$  auf:

$$\gamma_{io} = - \sum_{u=1}^v \gamma_{iu} c_{max}^u$$

Das setzt man in die Regression ein:

$$y_r = \alpha_1 x_{r1} + \dots + \sum_{u=1}^v \gamma_{iu} \left\{ \sum_{s \in U_r} (c_{rs}^u - c_{max}^u) \cdot z_{is} \right\} + \dots + \alpha_p x_{rp} + \epsilon_r \quad (11)$$

Die Ausdrücke in  $\{ \}$  sind von den unbekanntten Parametern unabhängige Regressoren. Man sieht sofort, daß die Wirkung von  $z_{is}$  auf  $y_r$  voraussetzungsgemäß gegen Null geht, wenn  $c_{rs}$  gegen  $c_{max}$  geht.

Durch dieses modifizierte Almon-Verfahren beseitigt man zwar die unerwünschte Sprungstelle in der Zugangsfunktion, alle anderen Kritikpunkte am Almon-Verfahren gelten jedoch unverändert.

Wie bereits angedeutet, läßt sich das Almon-Verfahren nicht problemlos auf Regressionen mit Durchschnittspotentialen übertragen; denn hier sind die Restriktionen (7) zu beachten, deren erste sich nur mit Glück und deren zweite sich im allgemeinen gar nicht, oder zumindest nicht genau erfüllen läßt, ohne die lineare Struktur des Regressionsansatzes aufzugeben.

Wir wollen kurz zeigen, auch welche Schwierigkeiten das Almon-Verfahren bei Durchschnittspotentialen führt. Offensichtlich können die Gewichtungsfunktionen nicht wie die Zugangsfunktion

in (8) als Polynome mit regionsunabhängigen Koeffizienten geschrieben werden, weil sich damit die Bedingungen  $\sum_{s \in U_r} w_r(\gamma_i, c_{rs}) = 1$  im allgemeinen nicht für alle  $r$  erfüllen lassen. Stattdessen könnten wir die Gewichtungsfunktionen

$$w_r(\beta_i, c_{rs}) = \left( \sum_{u=0}^v \beta_{iu} c_{rs}^u \right) \left( \sum_{s \in U_r} \sum_{u=0}^v \beta_{iu} c_{rs}^u \right)^{-1} \quad (12)$$

verwenden, vorausgesetzt, der Ausdruck  $\sum_{u=0}^v \beta_{iu} c_{rs}^u$  ist für  $0 \leq c_{rs} \leq c_{\max}$  nichtnegativ. In der Formel (12) kann man den Vektor  $\beta_i$  mit einem beliebigen Skalar multiplizieren, ohne  $w_r$  zu verändern. Wir schließen diesen Freiheitsgrad durch die Festlegung  $\beta_{i0} = 1$ . Damit bleiben nur noch  $\beta_{i1}$  bis  $\beta_{iv}$  als zu schätzende Parameter.

Wir setzen jetzt (12) für  $w_r$  in das Durchschnittspotential (6) und dieses dann für  $x_{ri}$  in die Regression (1) ein. Dann erhalten wir

$$Y_r = \alpha_1 x_{r1} + \dots + \sum_{u=0}^v \gamma_{iu} \left\{ \sum_{s \in U_r} c_{rs}^u z_{is} \right\} S_{ri}^{-1} + \dots + \alpha_p x_{rp} + \epsilon_r \quad (13)$$

als Pendant zu (9). Hier ist - wie gehabt -  $\gamma_{iu} = \alpha_i \beta_{iu}$ , also insbesondere  $\gamma_{i0} = \alpha_i$ . (13) unterscheidet sich von (9) durch die vom unbekanntem Parametervektor  $\beta_i$  abhängige Größe  $S_{ri}$ , die definiert ist als

$$S_{ri} = \sum_{s \in U_r} \sum_{u=0}^v \beta_{iu} c_{rs}^u \quad (14)$$

Fordern wir im modifizierten Almon-Verfahren, daß die Nebenbedingung  $w_r(\beta_i, c_{\max}) = 0$  für alle  $r$  erfüllt sein soll, ergibt sich als Pendant zu (11)

$$Y_r = \alpha_1 x_{r1} + \dots + \sum_{u=1}^v \gamma_{iu} \left\{ \sum_{s \in U_r} (c_{rs}^u - c_{\max}^u) z_{is} \right\} S_{ri}^{-1} + \dots + \alpha_p x_{rp} + \epsilon_r \quad (15)$$

Da  $S_{ri}$  von  $\beta_i$  abhängt und über die Regionen variiert, sind die

Ansätze (13) bzw. (15) nicht mehr linear in den Parametern. Als einfaches Schätzverfahren ließe sich folgendes iteratives Verfahren ausprobieren: Man beginnt mit irgendeiner Vorgabe für  $\beta_i$ , errechnet nach (14)  $S_{ri}$ , setzt dies in die Regression (13) bzw. (15) ein, schätzt  $\alpha$  und  $\gamma_{iu}$  und errechnet daraus eine Schätzung für  $\beta_i$  nach  $\hat{\beta}_{iu} = \hat{\gamma}_{iu} / \hat{\gamma}_{i0}$ , wobei im modifizierten Almon-Verfahren  $\hat{\gamma}_{i0}$  nach

$$\hat{\gamma}_{i0} = - \sum_{u=1}^v \hat{\gamma}_{iu} c_{\max}^u$$

zu errechnen ist. Die Schätzung setzt man dann wieder in (14) ein, usw., bis sich  $\hat{\beta}_i$  nicht mehr wesentlich ändert. Im modifizierten Almon-Verfahren wäre eine gute Vorgabe für  $\beta_i$  z.B.  $\beta_{i0} = 1$ ,  $\beta_{i1} = -c_{\max}^{-1}$  und  $\beta_{iu} = 0$  für  $u > 1$ .

Das Verfahren ist nur brauchbar, wenn erstens die Bedingung

$$\sum_{u=0}^v \hat{\beta}_{iu} c_{rs}^u > 0 \quad \text{für} \quad 0 < c_{rs} < c_{\max}$$

durch die Parameterschätzungen erfüllt wird und zweitens der ange deutete Iterationsprozeß konvergiert. Anderenfalls muß man zu den schwierigeren nichtlinearen Methoden übergehen.

### 3.2 Verfahren mit Regionsringen

In der Zeitreihenanalyse kann der zeitliche Abstand innerhalb eines vorgegebenen Intervalls nur eine beschränkte Zahl von Werten annehmen (1,2,3,...Perioden). Bei den hier betrachteten räumlichen Querschnitten gilt für die Entfernung entsprechendes nicht. Doch kann man ihnen eine ähnliche Struktur aufzwingen, indem man eine Entfernungsvariable konstruiert, die nur eine kleine Zahl verschiedener Werte annehmen kann. Die Werte seien mit  $d_k$ ,  $k = 0, \dots, l$  bezeichnet. Der Entfernung jeder Region  $r$  zu ihren Umgebungsregionen ist jeweils einer der Werte  $d_k$

zuzuordnen. Die Menge von Regionen, deren Entfernung zur Region r der Betrag  $d_k$  zugeordnet ist, heißt Ring k-ter Ordnung um r und wird mit  $R_{rk}$  symbolisiert. Meist wird man als Ring nullter Ordnung die Region r selbst definieren.

Es gibt verschiedene Möglichkeiten, Ringe abzugrenzen und entsprechende Distanzen festzulegen. Man kann z.B. alle in einem Untersuchungsraum jeweils zwischen den Regionen und Ihren Umgebungsregionen auftretenden Entfernungen zu einer vorgegebenen Zahl von Größenklassen gruppieren (nach Augenschein oder nach einem Clusterverfahren) und dann jede Entfernung durch das entsprechende Klassenmittel ersetzen. Ein anderes Verfahren richtet die Ringabgrenzung ohne Berücksichtigung der Entfernung allein nach der Nachbarschaft: Mit  $k=1$  beginnend definiert man als Ring k-ter Ordnung um r die Menge aller Regionen, die mit einer Region des Ringes (k-1)-ter Ordnung eine gemeinsame Grenze haben und nicht bereits zu einem Ring mit niedriger Ordnung als k gehören. Als  $R_{r0}$  definiert man die Region selbst. Als Entfernung zwischen Region r und Regionen im Ring  $R_{rk}$  legt man dann z.B. die Ringordnung selbst fest (also  $d_k = k$ ), oder den Mittelwert der Entfernungen jeweils zwischen den Regionen und den Regionen ihres Umgebungsringes k-ter Ordnung.

Für den Fall des kumulativen Potentials können wir jetzt die Regressionsgleichung in der Form

$$Y_r = \alpha_1 x_{r1} + \dots + \sum_{k=0}^{l_i} \tilde{\alpha}_{ik} \sum_{s \in R_{rk}} z_{is} + \dots + \alpha_p x_{rp} + \epsilon_r \quad (14a)$$

schreiben, wobei  $l_i$  die für Variable i berücksichtigte maximale Ringordnung ist und  $\tilde{\alpha}_{ik}$  definiert ist als

$$\tilde{\alpha}_{ik} = \alpha_i \cdot a_i(\beta_i, d_k).$$

Legen wir dem Verlauf der Zugangsfunktion keinerlei Restriktionen auf, dann können wir die  $\tilde{\alpha}_{ik}$  wie üblich als Regressionskoeffizienten des linearen Ansatzes (14a) schätzen. Dabei können sich natürlich wie im Almon-Verfahren unplausible Schätzungen, z.B. mit wechselndem Vorzeichen einstellen.

Man kann das Verfahren an dieser Stelle abbrechen und auf eine explizite Schätzung einer Zugangsfunktion verzichten. Man betrachtet dann die Schätzungen  $\tilde{\alpha}_{ik}$  als die mit  $\alpha_i$  multiplizierten Werte der Zugangsfunktion an den Stellen  $d_k$ . Wir können stattdessen aber die Schätzung der Regression auch als Vorstufe zur expliziten Schätzung einer Zugangsfunktion verstehen, die wir dann funktional zu spezifizieren haben. Bei der Spezifikation sind wir hier nicht mehr auf die Polynom-Form angewiesen, sondern wählen z.B. aufgrund theoretischer Vorstellungen positive monoton sinkende Funktionen, in denen der Parameter  $\beta_i$  ausdrückt, wie stark sich der Einfluß mit wachsender Distanz reduziert.

Die gebräuchlichsten Funktionen hierfür sind die Potenzfunktionen  $a_i(\beta_i, d_k) = d_k^{-\beta_i}$ ,  $0 \leq \beta_i$  ; (15a)

und die geometrisch fallende Funktion  $a_i(\beta_i, d_k) = \beta_i^{d_k/\bar{d}}$ ,  $0 \leq \beta_i \leq 1$ . (16)

$\bar{d}$  ist die "Einheitsdistanz", die vorzugeben ist. Für die Schätzergebnisse ist die Vorgabe der Einheitsdistanz insofern ohne Belang, als die Menge der Zugangskurven<sup>1</sup>, die durch (16) bei unterschiedlichen Parametervorgaben erzeugt werden können, durch die Vorgabe von  $\bar{d}$  nicht beeinflusst wird. Durch  $\bar{d}$  wird lediglich gesteuert, welche Punkte des Parameterintervalls [0,1] den verschiedenen Zugangskurven zugeordnet sind.

Die geometrisch fallende Funktion kann äquivalent auch als Exponentialfunktion

$$a_i(\beta_i^*, d_k) = \exp(-\beta_i^* d_k), \quad 0 \leq \beta_i^* \quad (17)$$

geschrieben werden mit  $\exp(-\beta_i^* \bar{d}_k) = \beta_i$ . Man beachte, daß die  $\beta_i$  in (15a) und (16) (bzw.  $\beta_i^*$  in (17)) unterschiedliche Bedeutung haben. Während  $\beta_i$  in (15a) dimensionslos ist, hat  $\beta_i^*$  in (17) die Dimension [km<sup>-1</sup>], wenn  $d_k$

<sup>1</sup> Die Zugangskurve zeigt die Abhängigkeit des Wertes der Zugangsfunktion von der Distanz für festes  $\beta_i$ .

in [km] gemessen wird, bzw. [Stunden<sup>-1</sup>], oder ähnlich, wenn  $d_k$  in [Stunden] Fahrzeit oder ähnlich gemessen wird.<sup>1</sup> Anschaulicher als  $\beta_i^*$  ist die Halbwertsdistanz zu interpretieren. Sie ist definiert als

$$\bar{d}_i = (\ln 2) / \beta_i^*$$

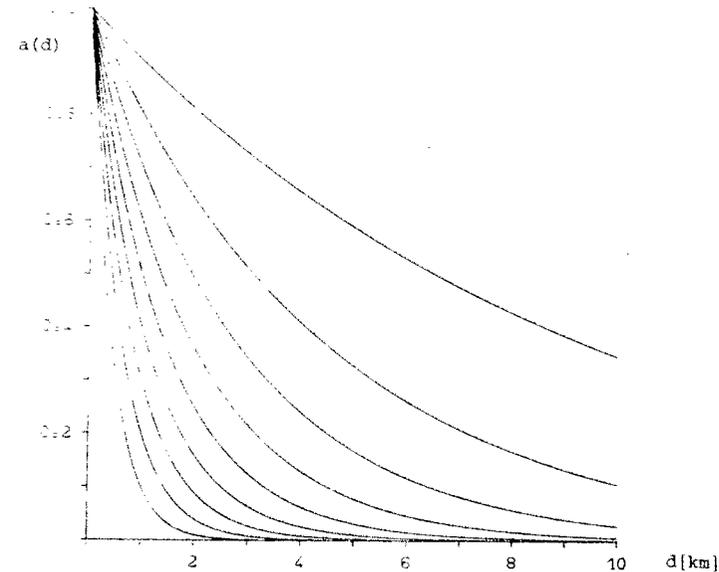
und gibt an, bei welcher Entfernung sich die Wirkung der Variable  $i$  auf die Hälfte reduziert.  $\beta_i$  in (16) ist der auf eine Einheitsdistanz bezogene Reduktionsfaktor. Er gibt an, um welchen Faktor sich die Wirkung reduziert, wenn man die Distanz um den Betrag einer Einheitsdistanz erhöht.

Abbildung 2 zeigt eine Schar Zugangskurven. Welcher Parameter  $\beta_i$  den einzelnen Kurven bei Vorgabe einer bestimmten Einheitsdistanz jeweils zugeordnet ist, kann man an der Ordinate ablesen, wenn man diese bis zur Stelle  $\bar{d}$  nach rechts verschiebt. Für  $\bar{d} = 1$  [Stunde] entsprechen den dargestellten Kurven die Parameter 0,1, 0,2, ... , 0,9. Je höher  $\bar{d}$  gewählt wird, desto stärker werden die Parameter von Kurven, die vergleichsweise starke Distanzwiderstände zum Ausdruck bringen, an den linken Rand der [0,1]-Skala gedrängt.

Die Potenzfunktion (15a) wurde früher am häufigsten verwendet, weil die Formel (3) für das kumulative Potential, die sich mit dieser Funktion ergibt, wenn man  $\beta_i = 1$  setzt, der Formel für das Gravitationspotential analog ist. Diese Analogie kann jedoch, wie heute wohl allgemein anerkannt ist, nicht als theoretische Begründung für die Verwendung dieser Funktionsform erhalten. Verglichen zum geometrisch fallenden Funktionsverlauf zeichnet sich die Potenzfunktion durch ein starkes Gefälle im Bereich kurzer und entsprechend schwächeres im Bereich langer Entfernungen aus. Vom praktischen Standpunkt her hat die Potenzfunktion den Nachteil, für  $d_k = 0$  den Wert unendlich anzunehmen. Wendet man also die Zugangsfunktion auch auf die Wirkung der Variablen der eigenen

<sup>1</sup> Es ist daher unsinnig, Parameterschätzungen für  $\beta_i$  in (17) anzugeben, ohne gleichzeitig die Dimension des Distanzmaßes zu nennen, wie es in der Literatur häufig geschieht.

Abbildung 2: Geometrisch fallende Zugangskurven



Region ( $z_{ir}$ ) an, muß man mit streng positiven intraregionalen Entfernungen arbeiten. Allerdings beeinflusst die gemessene intraregionale Entfernung ganz erheblich das relative Gewicht von  $z_{ir}$ , verglichen zu dem der  $z_{is}$ ,  $s \neq r$ . Dies ist unangenehm angesichts der Tatsache, daß der Festlegung einer intraregionalen Entfernung eine gewisse Willkür anhaftet. Es ist daher ratsam, die Potenzfunktion nur für den Zugang zu anderen Regionen, nicht für den zur eigenen Region, zu verwenden. Wir kommen darauf gleich noch zurück.

Wir nehmen nach diesem Exkurs die Frage wieder auf, wie im Anschluß an die beschriebene erste Schätzphase die  $\beta_i$ -Parameter der Zugangsfunktionen explizit zu schätzen sind. Man betrachtet die Schätzungen  $\hat{\alpha}_{ik}$  als Schätzungen der wahren Werte der Produkte  $\alpha_i \cdot a_i(\beta_i, d_k)$ . Die Schätzer  $\hat{\alpha}_{ik}$ ,  $k=0, \dots, l_i$  plottet man dann gegen die  $d_k$ ,  $k=0, \dots, l_i$ , und legt durch die dadurch entstehende Punktwolke die Funktion  $\alpha_i \cdot a_i(\beta_i, d_k)$ , indem man die Parameter  $\alpha_i$  und  $\beta_i$  so anpaßt, daß der Graph der Funktion sich an die Punktwolke optimal anpaßt.

Sowohl für die Potenz- als auch für die Exponentialfunktion liegt es nahe, dieses Anpassungsproblem dadurch zu lösen, daß man mit OLS die linearisierten Regressionen

$$\ln \hat{\alpha}_{ik} = \ln \alpha_i - \beta_i \ln d_k + \eta_k \quad \text{bzw.}$$

$$\hat{\alpha}_{ik} = \alpha_i - \beta_i^* d_k + \eta_k$$

schätzt. Das ist jedoch nicht ratsam, und zwar praktisch aus dem Grunde, daß dies  $\hat{\alpha}_{ik} > 0$  für alle  $k$  voraussetzt, was häufig nicht zutreffen wird, und theoretisch aus dem Grunde, daß hierdurch impliziert die stochastischen Abweichungen zwischen den wahren  $\hat{\alpha}_{ik}$  und ihren Schätzungen  $\hat{\alpha}_{ik}$  falsch spezifiziert werden.

Stattdessen ist die nichtlineare Regression

$$\hat{\alpha}_{ik} = \alpha_i \cdot a_i(\beta_i, d_k) + \mu_k \quad (18)$$

zu schätzen. Da die OLS-Schätzungen  $\hat{\alpha}_{ik}$  erwartungstreu sind (unter den üblichen Annahmen), hat  $\mu_k$  den Erwartungswert Null. Die Kovarianz-Matrix der  $\mu_k$ ,  $k=0, \dots, l_i$ , ist dieselbe wie die der  $\hat{\alpha}_{ik}$ ,  $k=0, \dots, l_i$ , und damit - bis auf einen geschätzten multiplikativen Faktor - bekannt. Wenn die Regression (14a) homoskedastisch ist und die Störungen unkorreliert sind, dann ist die Kovarianz-Matrix der  $\hat{\alpha}_{ik}$  die passende Submatrix der Matrix  $\sigma^2 (Z^T Z)^{-1}$ .<sup>1</sup> Hierbei bezeichnet  $\sigma^2$  die Varianz der Störungen in (14a),  $Z$  die Koeffizientenmatrix der Regression (14a) und  $T$  die Transposition. Zu jedem Parameter dieser Regression gehört eine Zeile bzw. Spalte der Matrix  $\sigma^2 (Z^T Z)^{-1}$ . Die Kovarianz  $\omega_{kk}$ , von  $\hat{\alpha}_{ik}$  und  $\hat{\alpha}_{ik}$ , ist dann das in der zu  $\hat{\alpha}_{ik}$  gehörenden Zeile und der zu  $\hat{\alpha}_{ik}$  gehörenden Spalte dieser Matrix stehende Element.

Da in (18) die Störungen korreliert sind, schätzt man  $\alpha_i$  und  $\beta_i$  nach dem nichtlinearen Minimum-Distance-Schätzer, d.h. man löst die quadratische Minimierungsaufgabe

<sup>1</sup> Ist allgemeiner die Kovarianz-Matrix der Störungen in (14a)  $\sigma^2 \Omega$ , dann ist  $\sigma^2 (Z^T Z)^{-1}$  durch  $\sigma^2 (Z^T \Omega^{-1} Z)^{-1}$  zu ersetzen. Vgl. etwa Schönfeld (1969, S. 138).

$$\min_{\alpha_i, \beta_i} \left\{ \sum_{k,k} \left[ \hat{\alpha}_{ik}^{-\alpha_i} \cdot a_i(\beta_i, d_k) \right] \omega_{kk} \cdot \left[ \hat{\alpha}_{ik}^{-\alpha_i} \cdot a_i(\beta_i, d_k) \right] \right\}.$$

Dabei sind die in (15a), (16) oder (17) geforderten Restriktionen einzuhalten. Das Minimum hinsichtlich  $\alpha_i$  für festes  $\beta_i$  läßt sich leicht explizit berechnen, indem man nach  $\alpha_i$  ableitet, das Ergebnis gleich Null setzt und nach  $\alpha_i$  auflöst. Das Resultat setzt man in die Zielfunktion ein und reduziert damit das Problem auf eine eindimensionale Minimierungsaufgabe, die sich numerisch problemlos iterativ lösen läßt.<sup>1</sup>

Verwendet man als Zugangsfunktion die Potenzfunktion (15), sollte man diese nur für  $k > 0$  an die  $\hat{\alpha}_{ik}$  anpassen. Als Schätzung für den Einfluß von  $z_{ir}$  auf das Wachstum verwendet man dann den in der ersten Phase geschätzten Koeffizienten  $\hat{\alpha}_{i0}$  anstatt des Funktionswertes der angepaßten Zugangsfunktion  $\hat{\alpha}_i d_0^{-\beta_i}$ . Dieses Vorgehen kann auch als ein Verfahren zur Schätzung der intraregionalen Distanz angesehen werden. Diese ergibt sich, wenn man die Gleichung  $\hat{\alpha}_{i0} = \hat{\alpha}_i d_0^{-\beta_i}$  nach  $\hat{d}_0$  auflöst.

Vor einem Resümee zu diesem Verfahren ist noch zu erwähnen, daß seine Anwendung auf die Schätzung von Distanzparametern in Durchschnittspotentialen auf dieselben Schwierigkeiten wie das Almon-Verfahren führt. Um die Restriktion (17) einzuhalten, wären jetzt die Zugangsfunktionen (15) oder (16) in Gewichtungsfunktionen zu verwandeln nach

$$w_r(\beta_i, d_k) = a_i(\beta_i, d_k) \left\{ \sum_k n_{rk} a_i(\beta_i, d_k) \right\}^{-1},$$

wobei  $n_{rk}$  die Anzahl von Regionen im Ring  $k$ -ter Ordnung um  $r$  bezeichnet. Aus der linearen Regression (14a) wird dann

$$y_r = \alpha_1 x_{r1} + \dots + \left\{ \sum_{k=0}^{l_i} \hat{\alpha}_{ik} \sum_{s \in R_{rk}} z_{is} \right\} s_{ri}^{-1} + \dots + \alpha_p x_{rp} + \varepsilon_r, \quad (19)$$

<sup>1</sup> Vgl. dazu z.B. die in der NAG-Programmbibliothek (NAG, 1982) angegebenen Methoden im Kapitel EO4.

wobei  $S_{ri}$  wieder eine a priori unbekannt, von  $\beta_i$  abhängige und über die Regionen  $r$  variierende Größe ist:

$$S_{ri} = \sum_k n_{rk} a_i(\beta_i, d_k)$$

Die Variation über  $r$  ergibt sich daraus, daß die Besetzungszahlen der verschiedenen Umgebungsringe sich von Region zu Region im allmeinen unterscheiden. Auch hier wäre daher mit einem analog zum Almon-Verfahren für Durchschnittspotentiale konstruierten Iterationsverfahren zu arbeiten, was hier nicht im einzelnen ausgeführt zu werden braucht. Allerdings geht die rechnerische Einfachheit - der einzige Vorteil des Ringbildungsverfahrens gegenüber den weiter unten dargestellten nichtlinearen Verfahren - jetzt weitgehend verloren, so daß das Festhalten an diesem Verfahren hier wohl nicht mehr angebracht ist.

Kommen wir zu einer abschließenden Beurteilung. Das Verfahren mit Regionsringen hat den Vorteil, zumindest in seiner Anwendung auf kumulative Potentiale rechnerisch ähnlich einfach wie das Almon-Verfahren zu sein, wenn auch in der zweiten Phase ein - allerdings simples - nichtlineares Minimierungsproblem zu lösen ist. Der Nachteil liegt in der mit der Bildung von Regionsringen einhergehenden Vergrößerung des Distanzmusters und der Notwendigkeit, die maximale Ringordnung  $l_1$  vorzugeben.

Der Anzahl  $l_1$  von Ringen, die man maximal berücksichtigen kann, sind enge Grenzen gesetzt, Jede Erhöhung von  $l_1$  um einen verringert die Zahl der in der Regression verbleibenden Freiheitsgrade um einen. Da in Regionsquerschnitten die Zahl der Beobachtungen meist nicht alzu üppig ist, ist man hier engen Schranken unterworfen. Sind die Ringe festgelegt, dann gibt man durch  $l_1$  praktisch einen Maximalradius vor, jenseits dessen eine Wirkung von Null unterstellt wird. Eine solche Vorgabe erweist sich im Nachhinein als problemlos, wenn sich aus der Schätzung ein vernachlässigbar kleiner Wert für  $a_i(\hat{\beta}_i, d_{l_1})$  ergibt. Anderenfalls muß man den Maximalradius weiter hinausschieben, was ohne Verlust von Freiheitsgraden nur möglich ist, indem man entsprechend breitere

Ringe bildet. Je breiter jedoch die Ringe sind, desto stärker wird das tatsächliche Distanzmuster durch die Ringbildung vergrößert, indem sehr unterschiedlichen Distanzen  $c_{rs}$  ein gemeinsamer Wert  $d_k$  zugeordnet wird.

Eingeschränkt ist die Brauchbarkeit des Verfahrens darüber hinaus dadurch, daß vorausgesetzt werden muß, daß man in der ersten Phase jeweils für ein  $i$  eine plausible Folge von  $\hat{\alpha}_{ik}$ -Schätzungen ( $k=0, \dots, l_1$ ) erhält, d.h. Schätzungen, die möglichst das Vorzeichen nicht wechseln und sich für wachsendes  $k$  monoton dem Wert Null nähern. Ist das nicht der Fall, wird man in der zweiten Phase nicht zu sinnvollen Schätzungen der Distanzparameter kommen. Allerdings kann dann auch die Frage aufgeworfen werden, ob sich nicht die in der Regression formulierten theoretischen Zusammenhänge überhaupt empirisch als unhaltbar erweisen.

Das Verfahren ist nach all dem Gesagten also nur bedingt brauchbar. Zumindest als Vorstufe zu den jetzt darzustellenden nichtlinearen Verfahren kann es aber in Betracht gezogen werden. Ein ungelöstes Problem dieser Verfahren ist nämlich die Festlegung von Startwerten im nichtlinearen Minimierungsprozeß. Als Startwerte kommen Schätzungen von Distanzparametern nach dem soeben dargestellten Verfahren in Frage.

### 3.3 Nichtlineare Verfahren

Es bleibt die Möglichkeit, die Regressionsparameter ( $\alpha$ -Parameter) und Distanzparameter ( $\beta$ -Parameter) durch die Anwendung nichtlinearer Verfahren zu schätzen. Konzeptionell ist das ein gradliniger Weg, in der Ausführung ist er jedoch aufwendig. Schreiben wir der Übersichtlichkeit halber den Ansatz noch einmal auf:

$$Y_r = \alpha_1 \epsilon_{r1}(\beta_1) + \dots + \alpha_q \epsilon_{rq}(\beta_q) + \alpha_{q+1} x_{r,q+1} + \dots + \alpha_p x_{rp} + \epsilon_p \quad (20)$$

Die ersten  $q$ -Regressoren sollen nichtlineare Funktionen jeweils

eines Distanzparameters (oder Parametervektors) sowie bekannter exogener Vektoren sein, die hier nicht explizit aufgeführt werden. Die restlichen Regressoren  $x_{r,q+1}$  bis  $x_{r,p}$  sollen bekannte deterministische Größen sein. Die  $\xi$ -Funktion können jetzt kumulative oder Durchschnittspotentiale, aber auch andere Arten von Accessibility-Indikatoren sein, von denen bereits die Rede war. Sie alle hängen davon ab, wie stark die den jeweiligen Indikatoren zugrundeliegenden Interaktionsphänomene durch die Distanz beeinflusst werden. Die Stärke des Distanzeinflusses wird durch den jeweiligen  $\beta$ -Parameter quantifiziert.

Für die weitere Darstellung schreiben wir (20) in der Form

$$y_r = f_r(\theta) + \varepsilon_r, \quad (21)$$

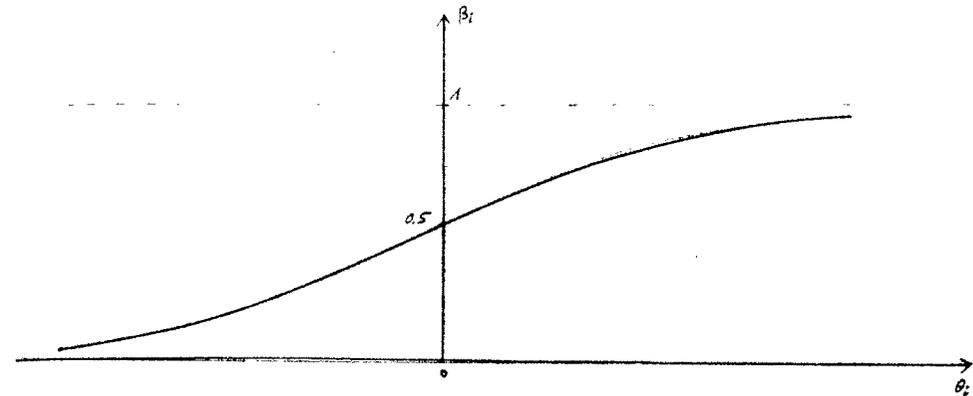
wobei  $\theta = (\beta_1, \dots, \beta_q, \alpha_1, \dots, \alpha_p)$  der unbekannte Parametervektor und  $f_r$  die durch (20) spezifizierte Funktion ist. Es sei  $m$  die Zahl der Beobachtungen und  $n$  die Dimension des Parameterraumes, also die Anzahl von Komponenten des Vektors  $\theta$ . Wenn die  $\varepsilon_{ri}$  jeweils nur von einem Parameter  $\beta_i$  (d.h. nicht von einem Parametervektor mit mehreren Komponenten) abhängen, dann ist  $n = q + p$ .

Wir wollen voraussetzen, die Parameter  $\theta_i, i=1, \dots, n$ , seien unbeschränkt. Nun haben wir oben Zugangsfunktionen mit beschränkten Parametern kennengelernt; z.B. ist  $\beta_i$  in (16) auf Werte von Null bis eins eingeschränkt. In diesem Falle schreibt man nicht  $\beta_i = \theta_i$ , sondern ersetzt  $\beta_i$  durch eine Funktion von  $\theta_i$ , die die ganze reelle Achse in den vorgegebenen Bereich transformiert. Man verwendet hier z.B. die Logit-Transformation

$$\beta_i = \frac{\exp(\theta_i)}{\exp(-\theta_i) + \exp(\theta_i)} \quad (22)$$

Für  $\theta_i = 0$  ist  $\beta_i = 0,5$  für  $\theta_i \rightarrow \infty$  geht  $\beta_i \rightarrow 1$  und für  $\theta_i \rightarrow -\infty$  geht  $\beta_i \rightarrow 0$ . Die Funktion hat die in Abbildung 3 gezeigte Gestalt.

Abbildung 3: Logit-Transformation



### 3.3.1 Kleinste-Quadrate-Prinzip

Es liegt nahe, den Parametervektor in (21) analog zum linearen Fall nach dem Kleinste-Quadrate-Prinzip (KQ-Prinzip) zu bestimmen, d.h. man sucht den Parametervektor  $\hat{\theta}$ , für den die Funktion

$$Q(\theta) = \sum_r (y_r^b - f_r(\theta))^2 \quad (22a)$$

ihre globales Minimum annimmt.  $y_r^b$  ist der beobachtete Wert der abhängigen Variable in Region  $r$ .

Bevor man versucht, dafür Lösungsverfahren anzugeben, muß man fragen, ob das KQ-Prinzip sich für den nichtlinearen Fall genau so begründen läßt wie für den linearen Fall. Die Antwort fällt leider negativ aus. Im allgemeinen gehen die wünschenswerten Eigenschaften, die KQ-Schätzer in linearen Modellen besitzen, im nichtlinearen Fall verloren. Es lassen sich lediglich einige wünschenswerte asymptotische Eigenschaften ableiten, deren praktischer Wert angesichts der üblichen Stichprobengrößen in regionalen Querschnitten allerdings fragwürdig ist.

Im linearen Fall erfährt das KQ-Prinzip bekanntlich seine Rechtfertigung daher, daß KQ-Schätzer die besten linearen unverzerrten

Schätzer (BLUE) sind, d.h. daß sie in der Klasse linearer und unverzerrter Schätzer diejenigen sind, die am wenigsten um den wahren Parametervektor streuen. Dies Resultat gilt unter der Voraussetzung, daß die  $\epsilon_r$  unkorreliert sind, gleiche Varianzen und Erwartungswerte von Null haben. Weitergehende Annahmen zum Verteilungsgesetz der Fehler, etwa die Normalverteilungsannahme, werden für dieses Resultat nicht benötigt. KQ-Schätzer sind darüber hinaus konsistent und asymptotisch normalverteilt, sofern bestimmte Regularitätsbedingungen beim Wachsen der Stichprobe über alle Grenzen eingehalten werden.

Nur die asymptotischen Eigenschaften lassen sich auf den nichtlinearen Fall übertragen. Nichtlineare KQ-Schätzer sind unter bestimmten Voraussetzungen konsistent und asymptotisch normalverteilt. Zu den Voraussetzungen gehört die Annahme, daß die  $f_r$  stetig und zweimal stetig differenzierbar sind, was in den von uns diskutierten Ansätzen der Fall ist; ferner die Annahme, daß die Störungen unabhängig identisch verteilt sind mit endlicher Varianz  $\sigma^2$  und Erwartungswerten von Null. Diese Annahme ist strenger als die für lineare OLS-Schätzungen getroffene Voraussetzung gleicher Varianzen und verschwindender Kovarianzen der Störungen. Weiterhin wird vorausgesetzt, daß der wahre Parameter  $\theta$  innerer Punkt in einem kompakten Teilbereich des Parameterraumes ist. Es ist also z.B. ausgeschlossen, daß  $\beta_i$  in der geometrisch fallenden Zugangsfunktion (16) den Wert 1 annimmt. Schließlich werden für die Herleitung der asymptotischen Resultate gewisse Regularitätsbedingungen verlangt, die beim hypothetisch betrachteten Wachstum der Stichprobengrößen über alle Grenzen einzuhalten sind. Hierzu sei auf die Literatur verwiesen (Jennrich, 1969; Malinvaud, 1970, Kap. 9; Schönfeld, 1971, S. 131).

Fordert man weitergehend, daß die Störungen normalverteilt sind, dann ist der aus der Minimierung von Q resultierende Parametervektor ein Maximum-Likelihood-Schätzer und besitzt damit die Eigenschaften von Maximum-Likelihood-Schätzern, insbesondere die der asymptotischen Effizienz.

Trotz der schwächeren theoretischen Begründung bleibt das Kleinste-Quadrate-Prinzip auch im nichtlinearen Fall ein anschaulich plausibles Schätzprinzip. Analog zur linearen GLS-Schätzung läßt sich auch ein nichtlinearer Minimum-Distanz-Schätzer entwickeln für den Fall einer nichtskalaren und bis auf einen Skalar bekannten Kovarianzmatrix der Residuen. Für die Praxis erscheint uns das aber von geringem Nutzen, da über Varianzen und Kovarianzen der Residuen meistens nichts bekannt ist und man daher als einfachste Annahme ein "weißes Rauschen" voraussetzen sollte. Wir bleiben daher beim Kleinste-Quadrate-Prinzip.

### 3.3.2 Vertrauensbereiche und Tests

Gelingt es uns, mit Hilfe von Verfahren, auf die im nächsten Abschnitt einzugehen ist, einen Vektor  $\hat{\theta}$  zu finden, der  $Q(\theta)$  global minimiert, stellt sich als nächstes die Frage nach der Verlässlichkeit der gewonnenen Schätzung.

Im linearen Fall ist das exakte Verteilungsgesetz der F-Statistik, auf die sich die bekannten Hypothesentests und Konstruktionen von Vertrauensbereichen zurückführen lassen, bekannt, wenn die Störungen normalverteilt sind. Selbst wenn keine Normalverteilung vorliegt, bleibt die F-Verteilung der F-Statistik für große Stichproben gültig. Wie zu vermuten, gilt entsprechendes im nichtlinearen Fall nicht mehr. Hier lassen sich nur noch approximative Tests und Vertrauensbereiche angeben, die sich auf Analogieüberlegungen zum linearen Fall oder auf asymptotische Überlegungen gründen. Die Brauchbarkeit für den Einzelfall ist fraglich, so daß wir uns hier auf unsicherem Boden bewegen.

Für die folgende Darstellung müssen verschiedene Matrizen eingeführt werden, die wir auch im nächsten Abschnitt noch benötigen. Die Jacobi-Matrix des Funktionenvektors  $f(\theta)$  ist die  $(m \times n)$ -Matrix  $J^T$  mit der Komponente  $\frac{\partial f_r}{\partial \theta_i}$  in Zeile r und Spalte i. Für die Regression (20) lautet die r-te Zeile dieser Matrix

$(\alpha_1 \xi'_{r1}(\beta_1), \dots, \alpha_q \xi'_{rq}(\beta_q), \xi_{r1}(\beta_1), \dots, \xi_{rq}(\beta_q), x_{r,q+1}, \dots, x_{rp})$ ,  
wobei  $\xi'_{ri}$  die Ableitung von  $\xi_{ri}$  nach  $\beta_i$  bezeichnet.<sup>1</sup>

Die Hesse-Matrix der Funktion  $f_r(\theta)$  ist die  $(n \times n)$ -Matrix  $G_r$  mit der Komponente

$\frac{\partial^2 f_r}{\partial \theta_i \partial \theta_j}$  in der Zeile i und Spalte j. für die Regression (20) sind nur wenige Komponenten dieser Matrix ungleich Null. Wenn die  $\xi$ -Funktionen jeweils nur von einem  $\beta$ -Parameter abhängen, dann stehen in den ersten q Feldern der Hauptdiagonale von  $G_r$  die Komponenten

$\alpha_i \xi''_{ri}(\beta_i)$ ,  $i = 1, \dots, q$ , wobei  $\xi''$  die zweite Ableitung nach dem Distanzparameter bezeichnet. Außerhalb der Hauptdiagonale stehen nur noch die Kreuzableitungen

$\frac{\partial^2 f_r}{\partial \alpha_i \partial \beta_i} = \xi'_{ri}(\beta_i)$  für alle i in der jeweils zu  $\alpha_i$  und  $\beta_i$  gehörenden Zeile und Spalte.

Schließlich benötigen wir noch die  $(n \times n)$ -Matrix H mit der Komponente

$\frac{1}{2} \frac{\partial^2 Q}{\partial \theta_i \partial \theta_j}$  in Zeile i und Spalte j, also die mit dem Faktor  $\frac{1}{2}$

multiplizierte Hesse-Matrix der Residuenquadratsumme  $Q(\theta)$ . Man rechnet leicht nach, daß gilt

$$H = J^T J - \frac{1}{2} \sum_r (y_r^b - f_r(\theta)) G_r \quad (23)$$

Für den Spezialfall der linearen Regression ist  $G_r = 0$  für alle r,

<sup>1</sup> Ist  $\beta_i$  beschränkt und wird daher für  $\beta_i$  eine Funktion von  $\theta_i$  eingesetzt, die die reelle Achse in den beschränkten Bereich transformiert (siehe z.B. Gleichung (22)), dann hat man statt der Ableitungen nach  $\beta_i$  die nach  $\theta_i$  zu bilden. Man hat also  $\xi'_{ri}$  zu ersetzen durch  $\xi'_{ri} \beta'_i$ , wobei  $\beta'_i$  die Ableitung von  $\beta_i$  nach  $\theta_i$  bezeichnet. Entsprechend ist  $\xi''_{ri}$  zu ersetzen durch  $\xi''_{ri} \beta''_i + \xi'_{ri}(\beta'_i)^2$ .

so daß hier  $H = J^T J$  gilt. Im nichtlinearen Fall gilt diese Gleichung, wenn alle Residuen verschwinden. Im allgemeinen ist im nichtlinearen Fall jedoch  $H \neq J^T J$ . Die Matrizen J,  $G_r$  und H hängen von  $\theta$  ab, was der Einfachheit halber nicht explizit hingeschrieben wird.

Hypothesentests und Konstruktionen von Konfidenzbereichen im linearen Regressionsmodell lassen sich auf die F-Statistik für den Test der Hypothese zurückführen, der wahre Wert des Vektors  $b = A\theta$  sei  $\bar{b}$ . A ist eine  $(n' \times m)$ -Matrix,  $n' < n$ , mit  $n'$  linear unabhängigen Zeilen. Die wichtigsten Sonderfälle ergeben sich, wenn A die Einheitsmatrix ist oder wenn A ein Zeilenvektor mit einer 1 in der i-ten Spalte und sonst Nullen ist. Im ersten Fall wird die Hypothese geprüft, der gesamte Vektor  $\theta$  nehme einen vorgegebenen Wert  $\bar{\theta}$  an, im zweiten Fall, die i-te Komponente von  $\theta$  nehme einen vorgegebenen Wert  $\bar{\theta}_i$  an, während die anderen Komponenten frei variieren. Die F-Statistik lautet

$$F = \frac{Q(\theta^*) - Q(\hat{\theta})}{Q(\hat{\theta})} \frac{m-n}{n'} \quad (24)$$

mit  $Q(\hat{\theta}) = \min_{\theta} Q(\theta)$  und  $Q(\theta^*) = \min_{\theta \text{ s.t. } A\theta = \bar{b}} Q(\theta)$ .

Unter  $H_0$ , d.h. bei Gültigkeit der Nullhypothese  $b = \bar{b}$ , ist F verteilt wie das zentrale  $F(n', m-n)$ . Aufbauend auf (24) entwickelt man als Konfidenzbereich für  $\hat{b} = A\hat{\theta}$  bei der Irrtumswahrscheinlichkeit  $\alpha$ :

$$K_e = \left\{ b: \frac{Q(\theta^*) - Q(\hat{\theta})}{Q(\hat{\theta})} \frac{m-n}{n'} < F_{\alpha}(n', m-n) \right\} \quad (25)$$

Im linearen Fall läßt sich  $K_e$  leicht explizit berechnen. Man erhält ein  $n'$ -dimensionales Ellipsoid mit  $\hat{b}$  als Mittelpunkt. Die Hypothese, das wahre b liege außerhalb von  $K_e$ , kann mit höchstens  $\alpha$ -prozentiger Wahrscheinlichkeit für den Fehler erster Art zurückgewiesen werden. F kann analog nach (24) auch im nichtlinearen Fall errechnet werden, wenn für  $Q(\theta)$  die Funktion (22a) eingesetzt wird (Vgl. Goldfeld und Quandt, 1972, S. 50-57). Doch hier gibt es zwei gewichtige Einschränkungen: Erstens ist im nichtlinearen

Fall  $F$  nicht exakt, sondern nur asymptotisch  $F$ -verteilt. Die Güte der  $F$ -Approximation für das Verteilungsgesetz von  $F$  in (24) für den konkreten Fall einer nur mäßig großen Stichprobe ist unbekannt. Zweitens ist die Berechnung von  $F$  aufwendig, da für jeden Test, d.h. üblicherweise für jeden einzelnen Parameter, ein nichtlineares Minimierungsproblem zu lösen ist. Das ist oft ein nicht vertretbarer Rechenaufwand.

Es gibt lediglich einen Sonderfall, für den man (24) einfach errechnen kann. Testet man nämlich die Hypothese, alle Distanzparameter nähmen gleichzeitig bestimmte Werte an (z.B. Werte, die ein Verschwinden jeglicher interregionaler Wechselwirkung ausdrücken), dann reduziert sich die Berechnung von  $Q(\hat{\theta}^*)$  auf ein lineares Regressionsproblem. Der Rechenaufwand ist also praktisch bedeutungslos.

Der durch (25) definierte Bereich  $K_e$  wird im nichtlinearen Fall als exakter Konfidenzbereich bezeichnet (daher der Index  $e$ ). Exakt ist daran allerdings nur die Menge der  $b$ , für die der Anstieg der Residuenquadrate über den Wert  $Q(\hat{\theta})$  einen bestimmten Prozentsatz nicht überschreitet. Die zugehörige Irrtumswahrscheinlichkeit  $\alpha$  ist nicht exakt.

Um den Rechenaufwand zu umgehen, verwendet man üblicherweise Approximationen für  $F$ , die auf der quadratischen Approximation für  $Q(\theta)$  an der Stelle  $\hat{\theta}$  basieren. Weil  $\hat{\theta}$  die Funktion  $Q(\theta)$  minimiert, verschwinden die partiellen Ableitungen von  $Q(\theta)$  an der Stelle  $\hat{\theta}$ , so daß man als Taylor-Approximation für  $Q(\theta)$  bei Abbruch nach dem quadratischen Glied erhält:

$$\hat{Q}(\theta) = (\Delta\theta)^T H (\Delta\theta) + Q(\hat{\theta}), \quad (26)$$

mit  $\Delta\theta = \theta - \hat{\theta}$ . Weil  $\hat{\theta}$  die Funktion  $Q(\theta)$  zumindest lokal minimiert, ist  $H$  positivdefinit und  $\hat{Q}$  streng konvex, sofern ein strenges lokales Minimum vorliegt. Bei einem schwachen lokalen Minimum ist  $H$  positivsemidefinit und  $\hat{Q}$  konvex.

Wollen wir jetzt  $F$  unter Verwendung der Approximation errechnen, ersetzen wir  $Q(\theta^*)$  in (24) durch

$$\hat{Q}(\hat{\theta}^*) = \min_{\theta} \hat{Q}(\theta) \quad \text{s.t. } \hat{b} = A\theta.$$

Wegen Konvexität erhält man mit der Methode der Lagrangen Multiplikatoren notwendige und hinreichende Minimumbedingungen. Die Lösung nach  $\hat{\theta}^*$  setzt man in (26) ein und das Resultat aus (26) dann für  $Q(\theta^*)$  in (24). Das führt nach etwas mühseliger Rechnerei auf das Resultat

$$F_a = \frac{(\Delta b)^T (A H^{-1} A^T)^{-1} (\Delta b)}{Q(\hat{\theta})} \cdot \frac{m-n}{n'}$$

mit  $\Delta b = \hat{b} - A\hat{\theta}$ . Für den Konfidenzbereich ergibt sich

$$K_a = \left\{ b: F_a \leq F_{\alpha}(n', m-n) \right\}.$$

Ist die Krümmung der  $f_r$ -Funktionen an der Stelle  $\hat{\theta}$  nicht allzu stark und/oder sind die Residuen im Optimum nicht allzu groß, dann erweist sich in der Praxis oft, daß der Ausdruck  $\frac{1}{2} \sum_r (y_r^b - f_r(\theta)) G_r$  im Verhältnis zu  $J^T J$  vernachlässigbar klein ist, so daß man  $H$  durch  $J^T J$  näherungsweise ersetzen kann. Wir schreiben daher für den approximativen Konfidenzbereich

$$K_a = \left\{ b: \frac{(\Delta b)^T (A M^{-1} A^T)^{-1} (\Delta b)}{Q(\hat{\theta})} \cdot \frac{m-n}{n'} \leq F_{\alpha}(n', m-n) \right\}, \quad (27)$$

wobei  $M = H$  im Falle exakter quadratischer Approximation und  $M = J^T J$  im Falle approximativer quadratischer Approximation. Im linearen Fall ist natürlich  $K_e = K_a$  und außerdem ist  $H = J^T J$  eine parameterunabhängige Matrix, so daß hier alle dargestellten Konzepte identisch sind. (27) ist dann die aus Lehrbüchern bekannte Formel für Konfidenzbereiche linearer Funktionen des Parametervektors. Die Tests und Konfidenzbereiche für den gesamten Vektor  $\theta$  und eine einzige Komponente  $\theta_i$  ergeben sich auch hier wieder, wenn man für  $A$  die Einheitsmatrix bzw. den Vektor mit der Eins als  $i$ -ter Komponente und sonst Nullen einsetzt. Man erhält im letzten Fall

$$F_a = (\hat{\theta}_i - \hat{\theta}_i)^2 / \hat{\sigma}_{\theta_i}^2 \quad \text{mit} \quad \hat{\sigma}_{\theta_i}^2 = \frac{Q(\hat{\theta})}{m-n} \cdot (M^{-1})_{ii}$$

wobei  $(M^{-1})_{ii}$  das i-te Diagonalelement von  $M^{-1}$  bezeichnet.  $F$  ist also das Quadrat der bekannten t-Statistik, wenn  $Q(\hat{\theta})/m-n$  als Schätzung der Kovarianzmatrix des geschätzten Parametervektors angesehen wird.

Bei den approximativen Konfidenzbereichen handelt es sich wieder um  $n'$ -dimensionale Ellipsoide. Um ihre Gestalt zu verdeutlichen, ist es hilfreich, die Matrix  $N = (A M^{-1} A^T)^{-1}$  zu zerlegen nach  $N = W L W^T$  mit einer Diagonalmatrix  $L$  und  $(n' \times n')$ -Matrizen  $W$  mit  $W^T W = W W^T = I$ .  $W$  ist die Matrix der orthonormalisierten Eigenvektoren und  $L$  die Diagonalmatrix der zugehörigen Eigenwerte, die hier alle reell und nichtnegativ und bei nichtsingulärer Matrix  $N$  sogar streng positiv sind (vgl. z.B. Bard, 1974, Appendix A-5). Die Spalten der Matrix  $W$  geben die Richtung der Hauptachsen des Konfidenzellipsoids an, und die Kehrwerte der Quadratwurzeln der Eigenwerte verhalten sich zueinander wie die Längen dieser Hauptachsen. Der Ellipsoid ist in Richtung derjenigen Spalten von  $W$  am längsten gestreckt, für die die kleinsten Eigenwerte auftreten. Im Grenzfall einer singulären Matrix wird mindestens ein Eigenwert Null. Es gibt daher mindestens eine entartete, d.h. unendlich lange Hauptachse. Variationen von  $\bar{b}$  entlang einer entarteten Hauptachse lassen  $Q(\hat{\theta})$  unverändert auf dem Niveau  $Q(\hat{\theta})$ .

Die von uns verwendeten Programme zur Schätzung nichtlinearer Regressionen stellen für den Sonderfall  $A = I$ , also  $N = M$ , die entsprechende Zerlegung

$$M = V \Lambda V^T$$

bereit. Sie gewährt nicht nur Einblick in die Gestalt des Konfidenzellipsoids des gesamten Parametervektors, sondern bildet auch einen guten Ausgangspunkt zur Berechnung von  $N^{-1}$  und damit zur Berechnung der geschätzten Kovarianzmatrix von  $\bar{b}$  nach

$$\text{Kov}(\hat{b}) = N^{-1} Q(\hat{\theta})/m-n. \quad (28)$$

Wir definieren die Matrix  $B = V^T A^T$ . Dann gilt  $A = B^T V^T$  bzw.  $A^T = V B$ . Wegen  $M^{-1} = V \Lambda^{-1} V^T$  folgt dann

$$N^{-1} = A M^{-1} A^T = B^T V^T V \Lambda^{-1} V^T V B = B^T \Lambda^{-1} B.$$

Für die Komponente  $(N^{-1})_{kl}$  in Zeile  $k$  und Spalte  $l$  der Matrix  $N^{-1}$  ergibt sich demnach in ausführlicher Schreibweise

$$(N^{-1})_{kl} = \sum_{i=1}^n \frac{B_{ik} B_{il}}{\lambda_i}, \quad (29)$$

wobei  $\lambda_i$  die  $i$ -te Diagonalkomponente von  $\Lambda$  und  $B_{ik}$  die Komponente in der Zeile  $i$  und Spalte  $k$  der Matrix  $B$  ist. Die praktische Bedeutung der Formel (29) liegt darin, daß sich  $(N^{-1})_{kl}$  nach dieser Formel auch dann berechnen läßt, wenn  $M$  singulär ist, aber die linearen Funktionen  $b_k$  und  $b_l$  des Parametervektors von der Singularität nicht betroffen sind.  $b_k$  heißt "von Singularität nicht betroffen", wenn für alle  $i$ , für die  $\lambda_i = 0$  ist, auch  $B_{ik} = 0$  gilt. Man berechnet dann  $(N^{-1})_{kl}$  einfach nach (29), indem man alle Summanden vernachlässigt, für die  $\lambda_i = 0$  ist (Vgl. Golub und Reinsch, 1971).

Insbesondere ist ein Parameter  $j$  von Singularität nicht betroffen, wenn in der Zeile  $j$  der Matrix  $V$  die Komponenten in allen Spalten  $i$  den Wert Null haben, für die  $\lambda_i = 0$  gilt.

Die Kovarianz zweier nicht von Singularität betroffener Parameter  $\theta_j$  und  $\theta_h$  ist dann

$$\text{Kov}(\theta_j, \theta_h) = \left( \sum_i \frac{V_{ji} V_{hi}}{\lambda_i} \right) Q(\hat{\theta})/m-n, \quad (30)$$

wobei wieder die Summanden weggelassen werden, für die  $\lambda_i = 0$  ist.

Bei dem von uns betrachteten Problem ist die Verwendung der Formeln, die auch bei Singularität die Berechnung von geschätzten Varianzen für die nicht betroffenen Parameter erlauben, von praktischer Bedeutung, weil Singularität z.B. daraus resultieren kann, daß bei bestimmten Distanzparametern z.B. zwei Potentiale miteinander hoch korreliert sind. Dann sind nur die zu diesen Potentiale gehörigen Parameter unsicher, während ansonsten möglicherweise brauchbare Ergebnisse vorliegen. Es ist umständlich, wenn unter Ausschaltung der Multikollinearität neu gerechnet werden muß, um diese Ergebnisse interpretieren zu können.

Es ist zum Abschluß dieses Abschnitts noch einmal hervorzuheben, daß wir uns wegen des approximativen Charakters der hier entwickelten Methoden auf weichem Grund bewegen. Dennoch haben die Teststatistiken heuristischen Wert und sagen meist immer noch mehr aus, als spontan aufgestellte ad hoc-Kriterien zur Beurteilung der Verlässlichkeit von Schätzergebnissen.

Häufig zieht man sich auf weniger tiefgründig hergeleitete Beurteilungskriterien zurück als die hier entwickelten; und dies hat durchaus seine Berechtigung, wenn damit bedeutende Vereinfachungen verbunden sind. Ohnehin genügen ja die von uns verwendeten Daten und geschätzten Modelle nicht den Anforderungen des statistischen Puritanismus, und die Rechenergebnisse versprechen bei weitem nicht die Exaktheit, die die theoretischen Begründungen der Methoden vortäuschen.

Von diesem eher handwerklichen Standpunkt aus sei schließlich erwähnt, daß wir oft mit der Angabe von Standardfehlern für die  $\alpha$ -Parameter, in denen die Regression linear ist, zufrieden sind. Solche lassen sich ermitteln, indem man die nichtlinearen Schätzungen für die Distanzparameter einfach in die  $\xi$ -Funktionen einsetzt, und die sich ergebenden Werte dieser Funktionen so wie die Werte bekannter deterministischer Regressoren behandelt. Man kann dann diese Regressoren in ein gängiges Programm für lineare Regression stecken und erhält alle gewünschten Resultate. Nach diesem Verfahren werden allerdings, wie aus (24) sofort ersichtlich ist, die Werte der F-Statistik überschätzt, also die Irrtumswahrscheinlichkeiten unterschätzt und die Vertrauensbereiche zu klein ausgewiesen. Man muß dem Verfahren allerdings zugute halten, daß derselbe Einwand gegen Tests von Parametern für Regressoren zu erheben ist, die nach den Kriterien optimaler Anpassung an die Beobachtungen aus einem Set potentieller Regressoren ausgewählt sind. Hier wird aber an der Verwendung gängiger Teststatistiken selten Anstoß genommen. Ihr heuristischer Wert gilt als allgemein akzeptiert!

### 3.3.3 Minimierung der Residuenquadrate

Das Grundsätzliche zur Frage, wie man das nichtlineare Kleinste-Quadrate-Problem numerisch löst, kann kurz gehalten werden, da dies Problem nicht nur in der Literatur ausführlich bearbeitet ist, sondern auch exzellentes Software zur Verfügung steht, das Selbstgetricktem unbedingt vorzuziehen ist angesichts einer Vielzahl diffiziler numerischer Hürden, die bei der Lösung des Problems zu überwinden sind.

Die verfügbaren Algorithmen konstruieren, ausgehend von einem Startwert  $\theta^0$ , eine für  $k \rightarrow \infty$  gegen ein lokales Minimum konvergierende Folge  $\{\theta^k\}$  mit

$$\theta^{k+1} = \theta^k + s^k r^k .$$

$k$  ist der Iterationsindex.  $s^k$ , ein Skalar, ist die Schrittlänge, und  $r^k$ , ein Vektor gleicher Dimension wie  $\theta$ , ist die Schrittrichtung. In jeder Iteration wird zuerst eine Schrittrichtung gewählt und dann in der festgelegten Richtung nach einem eindimensionalen Minimierungsverfahren die Schrittlänge so festgelegt, daß  $Q(\theta^{k+1})$  möglichst klein wird.

In den besten heute verfügbaren Verfahren basiert die Wahl der Schrittrichtung auf dem Newton-Verfahren, an dem allerdings eine Reihe von Modifikationen vorgenommen werden, um die bei einem reinen Newton-Verfahren nicht gewährleistete Konvergenz zu sichern. Die Newton-Richtung an der Stelle  $\theta^k$  ist die Richtung, in der von  $\theta^k$  aus gesehen das Minimum der quadratischen Approximation von  $Q(\theta)$ , entwickelt an der Stelle  $\theta^k$ , liegt.

Für unser Problem ergibt sich die Newton-Richtung als Lösung des Gleichungssystems

$$H r^k = J^T f , \tag{31}$$

wobei  $f$  den  $m$ -gliedrigen Spaltenvektor mit den Komponenten  $f_r(\theta)$  bezeichnet.  $H$ ,  $J$  und  $f$  werden an der Stelle  $\theta^k$  berechnet. (31) beschreibt allerdings nur die notwendige Optimumbedingung. Es wurde schon erwähnt, daß wir bei Vernachlässigung der zweiten Ableitungen  $H \approx J^T J$  setzen können, so daß wir statt (31)

$$J^T J r^k = J^T f \tag{32}$$

erhalten. Die Lösung von (32) heißt Gauss-Newton-Richtung (vgl. z.B. Gill und Murray, 1978).

In der Literatur wird am häufigsten auf modifizierte Gauss-Newton-Algorithmen verwiesen. Diese können sich jedoch als nachteilhaft gegenüber Newton-Verfahren erweisen, wenn  $J^T J$  eine schlechte Approximation für  $H$  ist, also insbesondere, wenn die Residuen groß sind. Angesichts der nur in der Größenordnung von 0.5 liegenden Bestimmtheitsmaße in Regionsquerschnitten ist davon auszugehen, daß die Residuen "groß" und die zweiten Ableitungen demnach nicht zu vernachlässigen sind. Zumindest, wenn Schritte in Gauss-Newton-Richtungen keine hinreichende Verbesserung pro Iteration ergeben, sollte daher auf die Newton-Richtung rekurriert werden.

Wir verwenden einen von Gill und Murray (1978) entwickelten Algorithmus, bei dem beide Richtungen in effizienter Weise kombiniert werden. Der Algorithmus nimmt außerdem an diesen Richtungen noch Modifikationen vor, um folgende Schwierigkeiten zu überwinden: Erstens kann die Newton-Richtung in die falsche Richtung, d.h. aufwärts weisen, wenn  $H$  nicht positivdefinit ist. Die Gauss-Newton-Richtung kann zwar nicht aufwärts, aber in eine Richtung weisen, die dem rechten Winkel zum steilsten Abstieg beliebig nahe kommt. Dies tritt ein, wenn  $J$  annähernd singulär wird. Man modifiziert daher bei der Lösung von (31) bzw. (32) die Matrizen  $H$  bzw.  $J^T$  derart, daß sie positivdefinit werden. Zweitens ist sicherzustellen, daß der Algorithmus auch dann voranschreitet, wenn die Matrizen  $H$  oder  $J$  (annähernd) singulär werden und sich die Inversen  $H^{-1}$  bzw.  $(J^T J)^{-1}$  nicht mehr (mit der notwendigen Genauigkeit) berechnen lassen. Drittens ist schließlich dafür Sorge zu tragen, daß der Algorithmus nicht gegen einen stationären Punkt konvergiert, der kein (lokales) Minimum ist.

Die Anwendung eines modifizierten Newton-Verfahrens erfordert nicht unbedingt, die ersten und zweiten Ableitungen der Regressionsfunktionen explizit zu berechnen. Stattdessen kann man mit Approximationsverfahren arbeiten, die aus wiederholten Berechnungen des Zielfunktionswertes auf den Gradienten und die Hesse-Matrix

der Zielfunktion schließen oder aus wiederholten Berechnungen des Zielfunktionswertes und des Gradienten auf die Hesse-Matrix schließen. Der Gradient läßt sich nach dem Verfahren der Sekantenapproximation annähern. Analog wird für die höheren Ableitungen verfahren. Alternativ wird zur Approximation der Hesse-Matrix häufig auch ein sogenanntes Quasi-Newton Updating-Verfahren angewandt, das sich in Beispielrechnungen von Gill und Murray (1978) für das nichtlineare Kleinste-Quadrate-Problem jedoch nicht so effizient zeigte, wie die guten Erfahrungen mit diesem Verfahren für generelle nichtlineare Optimierungsprobleme vermuten ließen.

Vom Anwendungsstandpunkt aus unterscheiden sich also die verfügbaren Algorithmen insbesondere danach, ob sie in den einzelnen Iterationen nur den Zielfunktionswert abfragen oder auch die ersten bzw. ersten und zweiten Ableitungen. Selbstverständlich müssen der Zielfunktionswert bzw. der Zielfunktionswert und der Gradient häufiger berechnet werden, wenn der Gradient und die Hesse-Matrix bzw. nur die Hesse-Matrix nicht explizit berechnet werden. Ob man einen Algorithmus wählt, der nur die Berechnung des Zielfunktionswertes, die Berechnung von Zielfunktionswert und Gradient oder die Berechnung von Zielfunktionswert, Gradient und Hesse-Matrix erfordert, hängt vom zusätzlichen Aufwand ab, der mit der expliziten Berechnung erster und zweiter Ableitungen verbunden ist. Im Falle einparametrischer Potentiale, die für die Praxis die größte Bedeutung haben, ist dieser Aufwand - verglichen zu einer wiederholten Berechnung des Zielfunktionswertes - gering, insbesondere dann, wenn man bei der Potentialberechnung interpoliert, wovon gleich noch die Rede sein wird. Daher ist ein Algorithmus zu wählen, bei dem für die Richtungswahl in jeder Iteration erste und zweite Ableitungen explizit errechnet werden. Ein entsprechendes FORTRAN-Unterprogramm steht in der NAG-Bibliothek (NAG, 1982) als Prozedur EO4HEF zur Verfügung. Es basiert auf den Ausführungen von Gill und Murray (1978) sowie den von diesen entwickelten Subroutinen der NPL<sup>1</sup>-Programm-bibliothek. Es fragt in jeder Iteration einmal die Matrizen  $J$  und  $G$  sowie ein- oder mehrmals den Vektor  $(y-f(\theta))$  ab.

-----  
<sup>1</sup> National Physical Laboratory, England.

Der große Rechenaufwand bei der Bereitstellung dieser Informationen ist mit der Berechnung der Potentiale sowie ihrer ersten und zweiten Ableitungen verbunden, die - wie unter 3.3.2 ausgeführt - in die Berechnung der Matrizen J und G eingehen. Hat man beispielsweise in einer Regression q kumulative Potentiale mit geometrisch fallenden Zugangsfunktionen und m Regionen<sup>1</sup>, dann sind zur Berechnung der Potentiale  $m^2q$  Exponentialfunktionen zu berechnen und ebensoviele Additionen durchzuführen. Für die Berechnung der ersten und zweiten Ableitungen sind jeweils noch einmal  $m^2q$  Multiplikationen und Additionen auszuführen. Das ist ein erheblicher zeitlicher Aufwand, der auch bei Problemen von moderatem Umfang schon zu unverträglich hohen Rechenzeiten führt. Auch der benötigte Speicherplatz ist sehr groß, weil man mindestens eine Distanzmatrix (bei je nach Potential unterschiedlichen Distanzmustern auch mehrere Distanzmatrizen) der Größe  $(m+1)m/2$  zur Verfügung haben muß.

Für den praktisch wichtigsten Fall ein-parametrischer Potentiale kann der Aufwand jedoch dadurch auf einen Bruchteil reduziert werden, daß man die Potentialfunktionen  $\varepsilon_{ri}(\beta_i)$  durch einfache Funktionsverläufe approximiert. Bevor wir dies im einzelnen erklären, sind jedoch einige Bemerkungen zur Parametrisierung der nichtlinearen Regression zu machen.

Die Art und Weise, wie die Regressionsfunktion (20) bzw. (21) parametrisiert wird, ist selbstverständlich nicht eindeutig. Ist z.B.  $\theta$  eine Funktion von  $\tilde{\theta}$ , d.h.  $\theta = g(\tilde{\theta})$ , wobei der Definitionsbereich von  $f$  der Wertebereich von  $g$  ist, können wir statt (21) schreiben

$$y_r = f_r(\tilde{\theta}) + \varepsilon_r = f_r(g(\tilde{\theta})) + \varepsilon_r, \quad (33)$$

so daß das Modell jetzt durch  $\tilde{\theta}$  anstatt durch  $\theta$  parametrisiert ist. Eine geschickte Wahl der Parametrisierung kann auf den Erfolg der

<sup>1</sup> Hier ist die Zahl von Regionen der Stichprobe und die Zahl von Regionen, auf der die Berechnung der Potentiale basiert, gleichgesetzt. In der Praxis können sie sich auch unterscheiden.

numerischen Optimierung großen Einfluß haben.

In der speziellen nichtlinearen Regression (20) ist auf eine für das numerische Rechnen nützliche Form der Parametrisierung der Potentialfunktionen zu achten. Hat man Potentiale nach (3) oder (6) berechnet, sollte man an diesen die folgenden Manipulationen vornehmen, die nicht das Modell selbst, sondern nur die Art seiner Parametrisierung berühren:

- (1) Man kann die Potentiale mit beliebigen, von  $\beta$  abhängigen und für alle r identischen Beträgen multiplizieren. Dies hat lediglich zur Folge, daß die  $\alpha$ -Parameter durch dieselben Beträge dividiert werden. Ist die Regression inhomogen, kann man darüber hinaus beliebige, von  $\beta$  abhängige und für alle Regionen identische Beträge zu den Potentialen hinzuaddieren. Dies hat lediglich zur Folge, daß dieselben Beträge, multipliziert mit den jeweils zugehörigen  $\alpha$ -Parametern, vom Achsenabschnitt subtrahiert werden. Durch Multiplikation mit, oder Subtraktion von geeigneten Beträgen, z.B. durch Standardisieren, sollte man daher den Einfluß von  $\beta$  auf das Niveau der Potentialfunktionen ausschalten.
- (2) Man kann die Parameter  $\beta_i$  in den Potentialen durch neue Parameter  $\tilde{\beta}_i$  ersetzen, indem man schreibt

$$\tilde{\varepsilon}_{ri}(\tilde{\beta}_i) = \varepsilon_{ri}(h_i(\tilde{\beta}_i)) \quad \text{mit} \quad \beta_i = h_i(\tilde{\beta}_i),$$

wobei  $h_i$  eine streng monotone, stetige und zweimal stetig differenzierbare Funktion sein soll. Für die im folgenden darzustellenden Methoden ist es hilfreich,  $h_i$  so zu wählen, daß  $\tilde{\beta}_i$  in einem vorgegebenen Intervall, z.B.  $[0,1]$ , liegt.<sup>1</sup> Mit anderen Worten soll  $h_i^{-1}$  den zulässigen Bereich für  $\beta_i$  in das Intervall  $[0,1]$  transformieren. Durch die Wahl von  $h_i$  ist es möglich, das Koordinatensystem, in das der Graph der Funktion  $\tilde{\varepsilon}_{ri}$  eingetragen wird, in Richtung der Abszisse in Teilbereichen des Intervalls  $[0,1]$  beliebig zu strecken oder zu stauchen. Durch Ausprobieren sollte man daher  $h_i$  so wählen, daß sich nicht für alle r z.B. scharfe Knickstellen

<sup>1</sup> Im Minimierungsprozeß wird das beschränkte  $\tilde{\beta}_i$  dann z.B. mit Hilfe der Logit-Transformation als Funktion des unbeschränkten Parameters  $\theta_i$  dargestellt. Vgl. oben, S. 23.

oder besonders steile Teilstücke der Funktion  $\xi_{ri}$  in bestimmten Teilstücken des Intervalls konzentrieren.<sup>1</sup>

Kommen wir jetzt zurück auf die Frage, wie sich die Potentiale und ihre ersten und zweiten Ableitungen vereinfacht berechnen lassen. Man geht dazu wie folgt vor. Zuerst berechnet man die auf dem Intervall  $[0,1]$  definierte Potentialfunktion<sup>2</sup>  $\xi_{ri}(\beta_i)$  an vorgegebenen "Stützstellen", z.B. an den 6 äquidistanten Stützstellen  $0, 0.2, 0.4, \dots, 1$ . Sodann legt man durch diese Punkte eine glatte Kurve, mit deren Hilfe sich sowohl die Funktionswerte zwischen den Stützstellen als auch die ersten und zweiten Ableitungen der Funktion approximieren lassen. Hierfür bietet sich das Verfahren der kubischen Spline-Approximation an. Nach diesem Verfahren wird der Kurvenverlauf zwischen jeweils zwei benachbarten Stützstellen durch eine kubische Funktion approximiert (Ahlberg und Walsh, 1967). Für den ersten Abschnitt  $[0, 0.2]$  beispielsweise ersetzt man  $\xi_{ri}(\beta_i)$  durch die Ersatzfunktion

$$\xi_{ri}^*(\beta_i) = a_{11} + a_{12} \beta_i + a_{13} \beta_i^2 + a_{14} \beta_i^3, \quad 0 < \beta_i < 0.2.$$

Jede derartige Ersatzfunktion hat 4 a-Koeffizienten. Sie sind so festzulegen, daß folgende Forderungen erfüllt sind:

- (1) An den Stützstellen stimmt der wahre Funktionswert mit dem der angrenzenden Ersatzfunktionen überein.
- (2) Die erste Ableitung der Approximation ist stetig, d.h. an den Stützstellen ist die Ableitung der links angrenzenden Ersatzfunktion gleich der der rechts angrenzenden Ersatzfunktion.
- (3) Entsprechendes gilt für die zweite Ableitung.

<sup>1</sup> Bei Verwendung der geometrisch fallenden Zugangsfunktion der Form (16) entspricht die Ersetzung einer Einheitsdistanz durch eine andere der Ersetzung einer Transformation durch eine andere. Beobachtet man hier beispielsweise besonders steil abfallende oder ansteigende Teilstücke in der Nähe von  $\beta_i=0$ , sollte man die Einheitsdistanz heruntersetzen, um die  $\beta_i$ -Achse im linken Bereich zu strecken und im rechten zu stauchen.

<sup>2</sup>  $\xi_{ri}(\beta_i)$  sei hier jetzt die auf dem Intervall  $[0,1]$  definierte und nach den unter (1) und (2) ausgeführten Gesichtspunkten modifizierte Potentialfunktion.

Diese Forderungen führen auf ein lineares Gleichungssystem zur Berechnung der a-Koeffizienten mit zwei Freiheitsgraden, die sich daraus ergeben, daß die Bedingungen (2) und (3) nicht für die Ränder gestellt werden können. Man fordert daher z.B. zusätzlich, daß an den Rändern die zweite Ableitung verschwinden soll. Kubische Spline-Approximationen lassen sich komfortabel z.B. mit der IMSL-Prozedur ICSICU berechnen (IMSL, 1982).

Wie man sieht, braucht man, um für ein gegebenes  $\beta_i$  die Potentiale und ihre ersten und zweiten Ableitungen auszurechnen, nur abzufragen, in welchem Stützbereich der Parameterwert liegt, und erhält dann mit wenigen Multiplikationen und Additionen die gewünschten Werte.

Die Interpolation lohnt sich natürlich nur dann, wenn die Zahl der Stützstellen, die man für eine befriedigende Approximation der Potentialfunktionen benötigt, deutlich kleiner ist als die Zahl der Funktionsabrufe während des Iterationsprozesses. Erfahrungsgemäß ist das der Fall. Man kommt z.B. für Bevölkerungspotentiale für Regierungsbezirke der Bundesrepublik leicht mit fünf Stützstellen aus, während man oft mehr als 50 Iterationen in der nicht-linearen Optimierung benötigt. Unter Vernachlässigung des geringen Aufwandes für die Errechnung von Funktionswerten und Ableitungen der Spline-Funktionen ergibt sich demnach ein Einsparungsfaktor von mindestens 1:10.

Es kommt hinzu, daß häufig viele verschiedene Regressionen zu rechnen sind, in denen überall dieselben Potentiale vorkommen. Beispielsweise schätzen Bröcker, Peschel und Reimers (1983) Wachstumsgleichungen für eine Vielzahl verschiedener Branchen, in denen immer wieder dieselben Potentiale als erklärende Variablen auftreten. Hier brauchen natürlich die Spline-Koeffizienten nicht jedesmal neu berechnet zu werden.

Zum Abschluß dieses Abschnitts muß noch das schwierige Problem der Startwertwahl behandelt werden. Die dargestellten Algorithmen schreiten von einem Startvektor  $\theta^0$  in einer Iterationsfolge zu einem lokalen Minimum von  $Q(\theta)$  vor. Da nicht gesichert ist, daß es nur

ein einziges lokales Minimum gibt, können unterschiedliche Startwerte zu unterschiedlichen Schätzungen führen, wobei keines der gefundenen lokalen Minima mit Sicherheit das globale ist. Man sollte daher versuchen, sich schon bei der Wahl des Startvektors möglichst an ein globales Minimum heranzutasten. Schwierig ist die Suche nach Startwerten nur hinsichtlich der  $\beta$ -Parameter. Hat man geeignete  $\beta$ -Startwerte gefunden, so vervollständigt man den Startvektor  $\theta^0$  durch diejenigen  $\alpha$ -Parameter, für die  $Q(\theta)$  mit den vorgegebenen  $\beta$ -Parametern sein globales Minimum annimmt. Diese sind nach OLS ja leicht auszurechnen.

Für die Wahl der  $\beta$ -Startparameter erweist sich die Beschränkung der  $\beta$ -Parameter auf ein Intervall wieder als hilfreich. Man kann dann z.B. in dem Würfel  $\{\beta_1, \dots, \beta_q: 0 \leq \beta_i \leq 1, i=1, \dots, q\}$  ein Gitter definieren und absuchen. Hat man etwa 4  $\beta_i$ -Parameter zu schätzen, dann kann man für alle  $\beta$ -Parameterkombinationen, die sich mit  $\beta_i = 0, \beta_i = \frac{1}{2}$  und  $\beta_i = 1$  für  $i=1, \dots, 4$  ergeben,  $Q(\theta)$  hinsichtlich der  $\alpha$ -Parameter minimieren. Das wären 81 lineare Regressionen, die sich schnell rechnen lassen. Man wählt als Startwert die Parameterkombination, bei der die Residuenquadratsumme am kleinsten ist. Derartige Gitter-Suchverfahren als Hilfsmittel globaler Optimierung lassen sich noch verfeinern. Auf jeden Fall kommen sie nur bei einer geringen Zahl von  $\beta$ -Parametern in Frage.

Als praktisch brauchbar hat sich auch ein Suchverfahren folgender Art erwiesen. Für die Regression (20) stellt man einen Set potentieller Regressoren auf, der zum einen aus den nicht potentialisierten Variablen  $x_{r,q+1}$  bis  $x_{rp}$  und zum anderen aus den Werten der Potentialfunktionen mit vorgegebenen Parametern besteht. Für jedes Potential  $i$  gibt man eine Reihe verschiedener Parameter (z.B.  $\beta_i = 0, 0.2, 0.4, \dots, 1$ ) vor, und die Werte der Potentialfunktionen für jeden dieser verschiedenen Parameter werden als Regressoren in den Set aufgenommen. Sodann wird eines der üblichen Regressorenauswahlverfahren<sup>1</sup> angewendet, um aus den verschiedenen Regressoren, die ein Potential mit verschiedenen Parametern

<sup>1</sup> Vgl. Draper und Smith, 1966, Kap. 6. Entsprechende Prozeduren stellen alle gebräuchlichen Regressionsprogramme bereit. Wir verwenden das Unterprogramm RLSTP der IMSL-Bibliothek (IMSL, 1982).

repräsentieren, jeweils denjenigen auszuwählen, bei dem das Bestimmtheitsmaß am größten ist. Der zugehörige  $\beta$ -Parameter wird dann als Startwert verwendet.

#### 4. Schlußfolgerung

In regionalen Querschnittsanalysen ist es meist unumgänglich, in irgendeiner Form zu berücksichtigen, daß der Wert der abhängigen Variable  $y_r$  in der Region  $r$  nicht nur von Variablen abhängt, die sich auf die eigene Region  $r$  beziehen, sondern "räumlich verzögert" auch von Variablen, die sich auf andere Regionen beziehen. Formal kann man dies durch Potentialisierung der erklärenden Variable bewerkstelligen.

Dabei stellt sich das Problem, gleichzeitig mit den Regressionsparametern auch die Distanzparameter der Potentiale zu schätzen, die die Stärke der "räumlichen Verzögerung" fernwirkender Variablen ausdrücken. Die Einstellung dieser Parameter hat erfahrungsgemäß einen erheblichen Einfluß auf die Schätzergebnisse. Unter bestimmten Bedingungen kann man dieses Schätzproblem als lineares Regressionsproblem formulieren. Dem Vorteil rechnerischer Einfachheit stehen jedoch erhebliche Nachteile gegenüber, die lineare Verfahren hier nur als bedingt tauglich erscheinen lassen.

Allgemein anwendbar auf das Problem ist die nichtlineare Kleinste-Quadrate-Schätzung. Sie liefert in der Praxis sinnvolle Schätzergebnisse, wenn auch einschränkend gesagt werden muß, daß über die statistischen Eigenschaften der Schätzer für mäßig große Stichproben nichts bekannt ist. Auch Vertrauensbereiche für die Parameterschätzungen lassen sich berechnen. Allerdings haben sie nur approximativen Charakter und - verglichen zur linearen Regression - nur eingeschränkte Aussagekraft. Vom Rechenaufwand her hält sich die nichtlineare Schätzung im akzeptablen Rahmen, sofern die Zahl der Distanzparameter nicht zu groß ist (Größenordnungen von 5 lassen sich gut rechnen) und die Potentiale einparametrisch sind. Voraussetzung dafür ist der Einsatz des dargestellten Interpolationsverfahrens zur Berechnung der Potentiale und ihrer Ableitungen.

Die hier dargestellten Methoden wurden in empirischen Untersuchungen über Determinanten des regionalen Wachstums in Skandinavien und der Bundesrepublik Deutschland mit Erfolg angewandt. Die Ergebnisse sind an anderer Stelle ausführlich erörtert (Bröcker, Peschel, Reimers, 1983).

#### Literaturverzeichnis

- Ahlberg, J.H., E.N. Nilson, J.L. Walsh (1967), *The Theory of Splines and Their Applications*, Academic Press, New York, San Francisco, London.
- Bard, J. (1974), *Nonlinear Parameter Estimation*, Academic Press, New York and London.
- Ben-Akiva, M., S.R. Lerman (1979), *Disaggregate Travel and Mobility-Choice Models and Measures of Accessibility*, in: J.A. Hansher and P.R. Stopher (eds), *Behavioural Travel Modelling*, Croom Helm, London.
- v. Böventer, E. (1975), *Regional Growth Theory, Urban Studies*, Vol. 12, S. 1-29.
- Bröcker, J. (1980), *An Application of Economic Interaction Models to the Analysis of Spatial Effects of Economic Integration, Environment and Planning A*, Vol. 12, S. 321-339.
- Derselbe (1984), *Interregionaler Handel und ökonomische Integration, empirische Modelle für westeuropäische Länder und Regionen*, V. Florentz, München.
- Bröcker, J., K. Peschel, W. Reimers (1983), *Regionales Wachstum und ökonomische Integration, eine empirische Modellstudie für Skandinavien und die Bundesrepublik Deutschland*, V. Florentz, München.
- Cliff, A.D., J.K. Ord (1973), *Spatial Autocorrelation*, Pion, London.
- Draper, N.R., H. Smith (1966), *Applied Regression Analysis*, Wiley, New York, hier zitiert nach der zweiten Auflage von 1981.
- Gill, P.E., and W. Murray (1978), *Algorithms for the Solution of the Nonlinear Least-Squares Problem*, *SIAM Journal of Numerical Analysis*, Vol. 15, S. 977-992.

- Goldfeld, S.M., and R.E. Quandt (1972), *Non-linear Methods in Econometrics*, North-Holland, Amsterdam.
- Golub, G.H., C. Reinsch (1971), *Singular Value Decomposition and Least Squares Solutions*, in: J.H. Wilkinson, C. Reinsch (eds), *Handbook for Automatic Computation, Vol. II, Linear Algebra*, Springer, Berlin.
- IMSL (1982), *IMSL-Reference Manual*, Ed. 9, Houston, Texas.
- Jennrich, R.I. (1969), *Asymptotic Properties of Non-Linear Least Squares Estimators*, *The Annals of Mathematical Statistics*, Vol. 40, S. 633-643.
- Johnston, J. (1963), *Econometric Methods*, McGraw-Hill, New York, hier zitiert nach der zweiten Auflage von 1972.
- Koll, R. (1979), *Regionales Wachstum, eine empirische Untersuchung seiner Bestimmungsfaktoren unter besonderer Berücksichtigung der Raumstruktur*, Minerva, München.
- Malinvaud, E. (1966), *Statistical Methods of Econometrics*, North Holland, Amsterdam, hier zitiert nach der dritten Auflage von 1980.
- NAG (1982), *Numerical Algorithms Group, Fortran Library Manual*, Mark 9, Oxford.
- Reimers, W. (1981), *Determinanten regionalen Wachstums, Stand der empirischen Forschung und ein neues ökonometrisches Modell für Skandinavien und die Bundesrepublik Deutschland*, V. Florentz, München.
- Reinhardt, F., H. Soeder (1977), *dtv-Atlas zur Mathematik, Tabellen und Texte, Band 2, Analysis und angewandte Mathematik*, Deutscher Taschenbuch Verlag, München, hier zitiert nach der vierten Auflage von 1982.
- Richardson, H.W. (1973), *Regional Growth Theory*, Macmillan, London and Basingstoke.
- Schönfeld, P. (1971), *Methoden der Ökonometrie, Band II*, Franz Vahlen, Berlin und Frankfurt.
- Williams, H.C.W.L. (1977), *On the Foundation of Travel Demand Models and Economic Evaluation Measures of User Benefit*, *Environment and Planning A*, Vol. 9, S. 285-344.