

Parameterschätzung in Regressionsmodellen mit räumlich korrelierten Störgrößen

Dissertation
zur Erlangung des Grades
eines Doktors der Naturwissenschaften
der Technischen Universität Dortmund

Der Fakultät Statistik der Technischen Universität Dortmund

vorgelegt von
Matthias Arnold
aus Dortmund

Dortmund, im März 2008

Inhaltsverzeichnis

1	Einleitung	1
2	Das lineare Regressionsmodell mit räumlich korrelierten Störgrößen	4
3	Parameterschätzung nach generalisierter Momentenmethode	12
3.1	Definition und Motivation	12
3.2	Ein alternativer Momentenmethodenschätzer	16
3.3	Konsistenz des alternativen Momentenmethodenschätzers . . .	20
4	Parameterschätzung nach Maximum-Likelihood-Methode	34
5	Erweiterungen des Modells	37
5.1	Residuen aus verallgemeinerter Kleinst-Quadrate-Regression .	37
5.2	Nichtlineare Regression	41
5.3	Andere Störgrößenstrukturen	43
6	Simulation	45
7	Anwendung auf Arbeitslosendaten	61
7.1	Die Daten	61
7.2	Das Modell	66
7.3	Ergebnisse	71
8	Zusammenfassung und Fazit	82
	Tabellenverzeichnis	85
	Abbildungsverzeichnis	85
	Literaturverzeichnis	86

1 Einleitung

In vielen Datensätzen stehen die Merkmalsträger in einer festen räumlichen Beziehung zueinander. Das ist beispielsweise in der Ökologie, der Epidemiologie oder auch im Bergbau der Fall. Die Merkmalsträger sind dann bestimmte Gebiete, in denen etwa Schadstoffgehalte, die Häufigkeit von Erkrankungen oder auch die Konzentration von Rohstoffen gemessen werden. Es liegt auf der Hand, dass diese Messungen in benachbarten Gebieten nicht unabhängig voneinander sind. So können Luftschadstoffe in einem Gebiet durch den Wind leicht in benachbarte Gebiete gelangen und Krankheiten können sich durch Ansteckung ausbreiten. Auch die Konzentrationen von Rohstoffen im Boden benachbarter Gebiete sind nicht unabhängig - sie hängen oft von den gleichen geologischen Konstellationen ab.

Bei der Modellierung solcher Abhängigkeiten ist ein wesentlicher Unterschied zu Zeitreihendaten zu berücksichtigen. Wie schon Whittle (1954) feststellt, ist bei Zeitreihen die Richtung der Abhängigkeit klar: Werte aus der Vergangenheit beeinflussen möglicherweise zukünftige Beobachtungen, aber niemals umgekehrt. Im Gegensatz dazu bestehen bei räumlich angeordneten Daten potentielle Abhängigkeiten in alle Richtungen.

Diese Arbeit betrachtet ein Modell räumlich korrelierter Beobachtungen. Dabei setzt sich eine Beobachtung aus einer Linearkombination der Beobachtungen in benachbarten Gebieten sowie einem gebietsspezifischen Term zusammen. Ein räumlicher Korrelationsparameter steuert die Stärke der Abhängigkeit zwischen den Gebieten.

Dieser Parameter kann nach der Maximum-Likelihood-Methode geschätzt werden, wobei meistens eine Normalverteilung der Beobachtungen angenommen wird. Da dieser Schätzer keine geschlossene Darstellung besitzt, müssen zur Berechnung numerische Optimierungsverfahren verwendet werden. Insbesondere bei großen Datensätzen können diese jedoch instabil sein. Als Alternative schlagen Kelejian und Prucha (1999) einen Schätzer nach verallgemeinerter Momentenmethode vor.

In den allermeisten Anwendungen werden für die einzelnen Gebiete verschiedene Erwartungswerte zugelassen, deren Abhängigkeit von Kovariablen in einem Regressionsmodell geschätzt wird. Die räumliche Korrelation überträgt sich dann auf die Störterme einer solchen Regressionsbeziehung.

Der Schätzer nach verallgemeinerter Momentenmethode verwendet theoretische Momente der Störgrößen und setzt diese den empirischen Momenten gleich. Da die Störgrößen jedoch nicht beobachtbar sind, müssen dabei die empirischen Momente der Regressionsresiduen verwendet werden. Obwohl der Unterschied zwischen Störgrößen und Residuen für wachsenden Stichprobenumfang verschwindet, gibt es für jeden endlichen Stichprobenumfang sehr wohl Unterschiede. Darum schlägt diese Arbeit eine Modifikation des Schätzers nach verallgemeinerter Momentenmethode vor. Da die empirischen Momente nur für die Residuen vorliegen, werden die theoretischen Momente auch für die Residuen berechnet.

Die Arbeit ist folgendermaßen strukturiert: Das nächste Kapitel stellt das lineare Regressionsmodell mit den oben beschriebenen räumlich korrelierten Störgrößen vor. Das dritte Kapitel behandelt die Parameterschätzung nach verallgemeinerter Momentenmethode. Nach einer kurzen Darstellung des Schätzers aus Kelejjan und Prucha (1999) präsentiert dieses Kapitel den modifizierten Schätzer sowie den Beweis seiner Konsistenz. Das folgende Kapitel 4 überträgt die Modifikation des Schätzers nach verallgemeinerter Momentenmethode auf den Maximum-Likelihood-Schätzer. Kapitel 5 diskutiert einige Erweiterungen der Schätzprozeduren, etwa auf nichtlineare Regressionen sowie verallgemeinerte Kleinst-Quadrate-Schätzung. Eine Simulationsstudie in Kapitel 6 vergleicht die finiten Eigenschaften des modifizierten Schätzers mit denen des ursprünglichen Schätzers für verschiedene Stichprobenumfänge. Das siebte Kapitel wendet den modifizierten Schätzer auf einen Datensatz zur Arbeitslosigkeit in Deutschland an. Hier zeigt sich, dass die Modellierung räumlicher Korrelationen nicht auf die zu Beginn genannten Bereiche wie Ökologie oder Epidemiologie beschränkt ist. Auch verschiede-

ne demographische Gruppen, die aus Kriterien wie Alter, Geschlecht oder Bildungsniveau gebildet werden, können als räumlich korreliert aufgefasst werden. Das letzte Kapitel schließlich fasst die Ergebnisse zusammen und nennt einige Ansatzpunkte für weitere Überlegungen.

2 Das lineare Regressionsmodell mit räumlich korrelierten Störgrößen

Im linearen Regressionsmodell

$$y_i = x_i^T \beta + u_i \quad (i = 1, \dots, n)$$

wird die Beobachtung y_i durch eine lineare Funktion der k im Vektor x_i zusammengefassten Ausprägungen der Kovariablen sowie den Störterm u_i erklärt. In Matrixschreibweise lautet das Modell

$$y = X\beta + u.$$

Der Vektor y besteht aus den n Beobachtungen der abhängigen Variablen, die $(n \times k)$ -Matrix X enthält die Ausprägungen der unabhängigen Variablen und die unbeobachtbaren Störgrößen u_i bilden den $(n \times 1)$ -Vektor u . Der unbekannte $(k \times 1)$ -Vektor β enthält die zu schätzenden Regressionsparameter. Der übliche Schätzer für β ist der Kleinst-Quadrate-Schätzer

$$\hat{\beta} = (X^T X)^{-1} X^T y,$$

wobei die Matrix X von vollem Spaltenrang sein muss, damit die Inverse definiert ist. Der Störgrößenvektor u wird geschätzt durch

$$\hat{u} = y - \hat{y} = y - X\hat{\beta} = \left[I - X (X^T X)^{-1} X^T \right] y =: My = Mu.$$

Unter üblichen Annahmen (vgl. Malinvaud (1970), S. 82ff) ist der KQ-Schätzer in gewissem Sinne optimal. Er hat unter allen linear erwartungstreuen Schätzern von β die kleinste Kovarianzmatrix bezüglich der Löwner-

Halbordnung. Falls u multivariat normalverteilt ist, gilt diese Optimalitätseigenschaft sogar innerhalb der größeren Klasse aller erwartungstreuen Schätzer.

Eine entscheidende Rolle spielt die Annahme, wonach die Störgrößen unkorreliert sind und alle die gleiche Varianz besitzen. Unter Normalverteilungsannahme können mit den Quantilen der t - bzw. F -Verteilung Hypothesentests für den unbekanntem Parametervektor β durchgeführt sowie entsprechende Konfidenzbereiche konstruiert werden.

Oft ist die Annahme einer skalaren Kovarianzmatrix für den Störgrößenvektor jedoch fraglich. In solchen Fällen ist der gewöhnliche KQ-Schätzer zwar immer noch erwartungstreu. Er verliert jedoch die oben angesprochene Eigenschaft der Effizienz. Auch die auf Quantilen der t - bzw. F -Verteilung basierenden Tests und Konfidenzbereiche verlieren ihre Gültigkeit.

Von diesen beiden Schwächen ist die erstere weniger schlimm. Der zweite Schwachpunkt hingegen ist gravierend, weil Tests und Konfidenzbereiche das vorgegebene Signifikanzniveau mitunter stark verfehlen. Eine Modellierung der Abhängigkeiten in den Störgrößen kann helfen, dieses Problem zu beheben.

Eine Möglichkeit dazu ist das Modell räumlich korrelierter Störgrößen, das von Whittle (1954) sowie Cliff und Ord (1973) vorgeschlagen wurde. Diese Art der Modellierung bietet sich an, wenn Beobachtungen y_i in verschiedenen, teilweise aneinandergrenzenden Gebieten erhoben wurden, etwa in Landkreisen oder Bundesländern. Auch die in Kapitel 7 analysierten gruppenspezifischen Arbeitslosenquoten können als räumlich korreliert aufgefasst werden - verschiedene demographische Gruppen übernehmen dann die Rolle der Gebiete.

Konkret sei im weiteren der Störgrößenvektor u räumlich korreliert gemäß

$$u = \rho W u + \varepsilon. \tag{1}$$

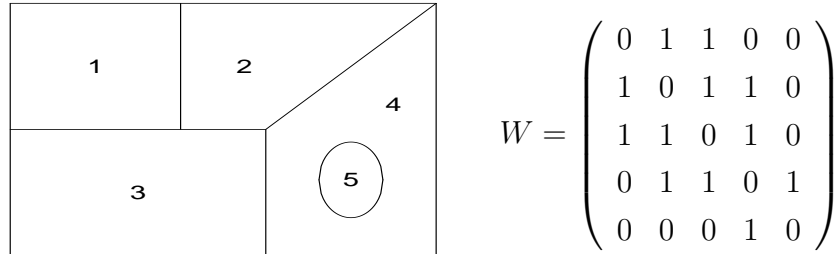


Abbildung 1: Beispiel für Verwandtschaftsmatrix

Der skalare Parameter ρ ist unbekannt. Die $(n \times n)$ -Verwandtschaftsmatrix W ist bekannt. Der Eintrag in Zeile i und Spalte j der Matrix W gibt das Ausmaß der möglichen Abhängigkeit zwischen u_i und u_j an. Für die Wahl der Einträge von W gibt es verschiedene Möglichkeiten. So kann etwa der Eintrag w_{ij} gleich Eins gesetzt werden, falls die Gebiete i und j eine gemeinsame Grenze haben. Abbildung 1 zeigt dazu ein kleines Beispiel. Das Gebiet 2 hat gemeinsame Grenzen zu den Gebieten 1, 3 und 4. Deshalb sind in der zweiten Zeile von W die Einträge 1, 3 und 4 von Null verschieden. Die zweite Zeile der Modellgleichung (1) lautet folglich

$$u_2 = \rho u_1 + \rho u_3 + \rho u_4 + \varepsilon_2.$$

Falls der Stichprobenumfang n einer Quadratzahl entspricht, bietet sich eine weitere Möglichkeit, die Matrix W zu spezifizieren. Die einzelnen Zellen können dann in einem Raster angeordnet werden; für $n = 16$ ergibt sich etwa

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

und die von Null verschiedenen Einträge in W orientieren sich an den Richtungen, in die Figuren auf einem Schachbrett ziehen können. Während der Turm nur entlang der Zeilen und Spalten zieht, darf die Dame auch die Diagonale benutzen. Im Gegensatz zu den Regeln beim Schach sind hier jedoch nur diejenigen Zellen verwandt, die mit einer Schrittlänge von Eins erreicht werden können. Für dieses Beispiel wären in der sechsten Zeile von W folgende Einträge von Null verschieden:

- nach dem „Turmkriterium“ die Einträge 2, 5, 7 und 10
- nach dem „Damenkriterium“ die Einträge 1, 2, 3, 5, 7, 9, 10 und 11.

Allgemeiner können die Einträge von W auch als Kehrwert eines Distanzmaßes, z. B. der räumlichen Distanz zwischen den Zentren der Gebiete i und j , gewählt werden. Wenn es sich bei den Untersuchungseinheiten um demographische Gruppen handelt, können diejenigen Gruppen als verwandt angesehen werden, bei denen bestimmte Merkmale übereinstimmen. In jedem dieser Fälle können die Zeilensummen der Matrix W anschließend noch auf Eins standardisiert werden. Die Standardisierung erfolgt über die Zeilen und nicht etwa die Spalten der Matrix W , weil dadurch die Modellgleichung (1) für die i -te Komponente von u ,

$$u_i = \rho \sum_{j=1}^n w_{ij} u_j + \varepsilon_i,$$

inhaltlich erhalten bleibt: Die Division der Einträge in der i -ten Zeile von W durch die Zeilensumme wird durch den Parameter ρ aufgefangen, der dadurch

zusätzlich besser interpretiert werden kann: Wegen $\sum_{j=1}^n w_{ij} = 1$ gibt ρ an, wie stark der Einfluss der benachbarten u_j auf u_i ist.

Wird die resultierende Kovarianzstruktur der Störgrößen nicht berücksichtigt und zur Schätzung von β der gewöhnliche KQ-Schätzer herangezogen, so hat dies die bereits oben erwähnten Konsequenzen: Der gewöhnliche KQ-Schätzer ist nicht effizient und Tests für den Parametervektor β verlieren ihre Gültigkeit. Krämer und Donninger (1987) vergleichen dazu die Schätzeigenschaften des gewöhnlichen mit denen des (effizienten) verallgemeinerten KQ-Schätzers. Es stellt sich heraus, dass die relative Effizienz beider Verfahren sehr stark von den Matrizen X und W abhängt. Diese Abhängigkeit von X und W findet Krämer (2003) auch für die Macht des F -Tests auf einen Teilvektor von β . Für manche ρ kann die Macht beliebig klein werden. Bei zeilenstandardisierter Verwandtschaftsmatrix W ist dies der Fall, wenn sich ρ dem Wert Eins annähert. Krämer (2005) zeigt weiter, dass zu jeder Verwandtschaftsmatrix W und jedem Stichprobenumfang n eine Modellmatrix X existiert, so dass dieses Problem einer verschwindenden Macht auch bei einem Test auf ρ selbst auftritt.

Formal seien für jeden Stichprobenumfang n die folgenden Annahmen über W und ε erfüllt:

Annahme 1 a) *Die Diagonalelemente von W sind gleich Null.*

b) *Sämtliche Einträge von W sind nicht negativ, $w_{ij} \geq 0 \forall i, j = 1, \dots, n$.*

c) *Die Zeilensummen von W sind gleich Eins, $\sum_{j=1}^n w_{ij} = 1 \forall i = 1, \dots, n$.*

d) *Die Spaltensummen von W sind beschränkt, $\sum_{i=1}^n w_{ij} \leq c_w \forall j = 1, \dots, n$ mit $0 < c_w < \infty$.*

e) $|\rho| < 1$.

Annahme 2 a) Die Innovationen $\varepsilon_1, \dots, \varepsilon_n$ sind unabhängig und identisch verteilt mit Erwartungswert Null und Varianz σ^2 , wobei die Varianz durch eine Konstante b beschränkt ist, $0 < \sigma^2 < b$ mit $b < \infty$.

b) $E(\varepsilon_i^4) < \infty$.

Unter diesen Annahmen ist die Matrix $(I - \rho W)$ invertierbar:

Satz 2.1 Für die Verwandtschaftsmatrix W seien die Annahmen 1(b), 1(d) und 1(e) erfüllt. Dann ist die Matrix $I - \rho W$ invertierbar.

Beweis: Es genügt zu zeigen, dass die Eigenwerte von $I - \rho W$ positiv sind. Sei also λ ein beliebiger Eigenwert von $I - \rho W$. Dann gibt es einen n -dimensionalen Vektor $x \neq 0$, so dass gilt:

$$\begin{aligned}
 (I - \rho W)x &= \lambda x \\
 \Rightarrow (1 - \lambda)x &= \rho Wx \\
 \Rightarrow \|(1 - \lambda)x\|_\infty &= \|\rho Wx\|_\infty \\
 \Rightarrow |1 - \lambda| \|x\|_\infty &= |\rho| \|Wx\|_\infty \leq |\rho| \|W\|_\infty \|x\|_\infty < 1 \cdot 1 \cdot \|x\|_\infty \\
 \Rightarrow |1 - \lambda| &< 1 \\
 \Rightarrow \lambda &> 0.
 \end{aligned}$$

Dabei ist $\|x\|_\infty = \max_i \sum_{i=1}^n |x_i|$ die Maximumnorm des Vektors x und $\|W\|_\infty = \max_i \sum_{j=1}^n |w_{ij}|$ die durch die Maximumnorm induzierte Zeilensummennorm der Matrix W . \square

Damit lautet wegen

$$\begin{aligned}
u &= \rho W u + \varepsilon \\
\Leftrightarrow \varepsilon &= (I - \rho W)u \\
\Leftrightarrow u &= (I - \rho W)^{-1}\varepsilon
\end{aligned}$$

die Kovarianzmatrix des Störgrößenvektors

$$Cov(u) = \sigma^2(I - \rho W)^{-1}(I - \rho W^T)^{-1} =: \sigma^2 P P^T. \quad (2)$$

Die beiden folgenden technischen Annahmen werden für die Konsistenz der in Kapitel 3 vorgeschlagenen Schätzer für ρ und σ^2 nach der verallgemeinerten Momentenmethode benötigt.

Annahme 3 Die Zeilen- und Spaltensummen der Beträge der Einträge der Matrix $P = (I - \rho W)^{-1}$ sind beschränkt, $\sum_{j=1}^n |p_{ij}| \leq c_p \forall i = 1, \dots, n$, $\sum_{i=1}^n |p_{ij}| \leq c_p \forall j = 1, \dots, n$ mit $0 < c_p < \infty$.

Annahme 4 Für die Einträge des Störgrößenvektors u sowie des Vektors der Regressionsresiduen $\hat{u} = Mu$ gilt folgendes: Es existieren Folgen von Zufallsvektoren d_i und Δ mit $\frac{1}{n} \sum_{i=1}^n \|d_i\|^{2+\delta} = O_P(1)$ für ein $\delta > 0$ sowie $\sqrt{n} \|\Delta\| = O_P(1)$, so dass $|u_i - \hat{u}_i| \leq \|d_i\| \|\Delta\|$ für alle $i = 1, \dots, n$ gilt.

Die Zufallsvektoren d_i und Δ dürfen - wie auch die Verwandtschaftsmatrix W - vom Folgenindex n abhängen. Auf eine zusätzliche Indizierung der Größen wird jedoch verzichtet, um die Lesbarkeit zu verbessern. Annahme 4 ist beispielsweise erfüllt, falls die Matrix X nichtstochastisch ist, ihre Zeilen keinen Trend aufweisen und die Matrix $\frac{1}{n} X^T X$ für $n \rightarrow \infty$ gegen eine feste Matrix Q konvergiert, denn dann erfüllt

$$u_i - \hat{u}_i = x_i^T (\hat{\beta} - \beta)$$

die Bedingungen aus Annahme 4: In diesem Fall gilt $\frac{1}{n} \sum_{i=1}^n \|x_i\|^{2+\delta} = O(1)$ für beliebiges $\delta > 0$, so dass die x_i die Rolle der d_i übernehmen. Für den Vektor $\hat{\beta} - \beta$ ist

$$\mathbb{E} \left(\sqrt{n} \left(\hat{\beta} - \beta \right) \right) = 0$$

sowie

$$\text{Cov} \left(\sqrt{n} \left(\hat{\beta} - \beta \right) \right) = \sigma^2 \left(\frac{1}{n} X^T X \right)^{-1} \xrightarrow{n \rightarrow \infty} \sigma^2 Q.$$

Die k Komponenten von $\hat{\beta} - \beta$ sind somit nach Multiplikation mit \sqrt{n} jeweils $O_P(1)$. Diese Eigenschaft überträgt sich auf den Gesamtvektor $\|\hat{\beta} - \beta\|$, so dass er die Forderung an Δ in der Annahme erfüllt.

3 Parameterschätzung nach generalisierter Momentenmethode

3.1 Definition und Motivation

Zur Schätzung von ρ und σ^2 wird oft die Maximum-Likelihood-Methode verwendet. Wie Kelejian und Prucha (1999) ausführen, kann die Berechnung des ML-Schätzers jedoch numerisch aufwändig sein. Der Grund liegt darin, dass in jedem Iterationsschritt die Determinante der Matrix $I - \rho W$ berechnet werden muss.

Als Alternative bietet sich ein Schätzer nach der generalisierten Momentenmethode an. Ein solcher Schätzer wird von Kelejian und Prucha (1999) vorgeschlagen, wobei die Momente

$$\mathbb{E} \left(\frac{1}{n} \varepsilon^T \varepsilon \right) = \sigma^2, \quad (3)$$

$$\mathbb{E} \left(\frac{1}{n} \varepsilon^T W^T W \varepsilon \right) = \frac{\sigma^2}{n} \text{tr}(W^T W), \quad (4)$$

$$\mathbb{E} \left(\frac{1}{n} \varepsilon^T W^T \varepsilon \right) = 0 \quad (5)$$

verwendet werden. Die Momente ergeben sich aus der für jede $(n \times n)$ -Matrix A gültigen Beziehung

$$\mathbb{E} (\varepsilon^T A \varepsilon) = \text{tr}[A \text{Cov}(\varepsilon)] + \mathbb{E}(\varepsilon)^T A \mathbb{E}(\varepsilon)$$

(vgl. Theorem 1.5 auf Seite 9 in Seber und Lee (2003)), denn nach Annahme 2 gilt hier $\mathbb{E}(\varepsilon) = 0$ und $\text{Cov}(\varepsilon) = \sigma^2 I$. Weiter sind nach Annahme 1(a) die Hauptdiagonalelemente und somit auch die Spur der Matrix W gleich Null. Die aus Gleichung (1) gewonnene Darstellung für ε ,

$$\varepsilon = u - \rho Wu,$$

wird in die theoretischen Momente eingesetzt und diese mit den empirischen gleichgesetzt. Dies führt auf das Gleichungssystem

$$G \begin{pmatrix} \rho \\ \rho^2 \\ \sigma^2 \end{pmatrix} - g =: v(\rho, \sigma^2),$$

mit

$$G = \begin{pmatrix} \frac{2}{n} \hat{u}^T W \hat{u} & -\frac{1}{n} \hat{u}^T W^T W \hat{u} & 1 \\ \frac{2}{n} \hat{u}^T W^T W W \hat{u} & -\frac{1}{n} \hat{u}^T W^T W^T W W \hat{u} & \frac{1}{n} \text{tr}(W^T W) \\ \frac{1}{n} \hat{u}^T [W + W^T] W \hat{u} & -\frac{1}{n} \hat{u}^T W^T W W \hat{u} & 0 \end{pmatrix} \quad (6)$$

und

$$g = \left(\frac{1}{n} \hat{u}^T \hat{u}, \frac{1}{n} \hat{u}^T W^T W \hat{u}, \frac{1}{n} \hat{u}^T W \hat{u} \right)^T. \quad (7)$$

Der Vektor $v(\rho, \sigma^2)$ ist die Differenz des Vektors g von dem Produkt aus der (3×3) -Matrix G und dem Spaltenvektor $(\rho, \rho^2, \sigma^2)^T$. Er gibt also in Abhängigkeit von ρ und σ^2 an, wie groß der Differenzen der theoretischen Momente von den empirischen Momenten sind. Folglich definieren Kelejjan und Prucha (1999) einen Schätzer für ρ und σ^2 durch

$$(\hat{\rho}_{NLS}, \hat{\sigma}_{NLS}^2) = \text{argmin}\{v(\rho, \sigma^2)^T v(\rho, \sigma^2) : \rho \in [-a, a], \sigma^2 \in [0, b]\}, \quad (8)$$

wobei $a \geq 1$ und $b < \infty$. In (6) und (7) ersetzen die beobachtbaren Regressionsresiduen \hat{u} die unbeobachtbaren Störgrößen u . Die Berechnung dieses Schätzers nach generalisierter Momentenmethode ist wesentlich weniger

aufwändig als die Berechnung des Maximum-Likelihood-Schätzers, weil hier lediglich die Norm eines Vektors der Länge drei, $v(\rho, \sigma^2)$, durch Wahl der beiden Parameter ρ und σ^2 minimiert wird. Die einmalige Bestimmung von $v(\rho, \sigma^2)$ aus der Matrix G sowie dem Vektor g stellt ebenfalls kein Problem dar, da die hierzu benötigten Quadratformen mit deutlich geringerem Aufwand berechnet werden können als die beim Maximum-Likelihood-Schätzer in jedem Iterationsschritt benötigte Determinante der Matrix $I - \rho W$.

Dieser Schätzer nach generalisierter Momentenmethode wurde in der Literatur bereits auf verschiedene Datensätze angewendet. So analysiert etwa Tingvall (2004) die Frage, ob der seit dem zweiten Weltkrieg fortschreitende Abbau von Handelsschranken innerhalb der Länder der heutigen europäischen Union zu einer verstärkten Spezialisierung der Volkswirtschaften in den einzelnen Mitgliedsstaaten geführt hat. Er findet heraus, dass dies nur für kapitalintensive Industrien der Fall ist, wohingegen kein allgemeiner Trend zu einer verstärkten Spezialisierung vorliegt. Garrett und Wheelock (2006) untersuchen die Veränderung des Pro-Kopf-Einkommens während der großen Wirtschaftskrise in den dreißiger Jahren des letzten Jahrhunderts für die Bundesstaaten der USA. Sie stellen fest, dass die Unterschiede im Pro-Kopf-Einkommen zwischen den Bundesstaaten während der Abschwungphase zunahm, um in der folgenden Aufschwungphase wieder abzunehmen: In der Krise gingen die Pro-Kopf-Einkommen in Staaten mit ohnehin geringerem Einkommen prozentual stärker zurück als in Staaten mit hohem Einkommen; dafür profitierten die Staaten mit geringerem Einkommen überproportional von dem anschließenden Aufschwung. Bell und Bockstael (2000) wenden den Schätzer auf einen Mikrodatsatz zu Immobilienpreisen in Maryland, USA, an. Sie vergleichen die Ergebnisse des Momentenmethodenschätzers mit denen des herkömmlichen Maximum-Likelihood-Schätzers und kommen zu dem Schluss, dass beide Verfahren im wesentlichen die gleichen Resultate liefern. Aufgrund des deutlich reduzierten Rechenaufwands empfehlen sie die Anwendung des Momentenmethodenschätzers

insbesondere bei großen Datensätzen. Schlenker et al. (2006) untersuchen die Auswirkungen der globalen Klimaerwärmung auf die US-amerikanische Landwirtschaft und kommen zu dem Ergebnis, dass leichte Verbesserungen in den nördlich gelegenen Landesteilen durch stärkere Verschlechterungen im Süden überkompensiert werden, so dass der Gesamteffekt einer Klimaerwärmung auf die Landwirtschaft in den USA negativ ist. Anselin et al. (2004) wenden den Schätzer ebenfalls auf einen landwirtschaftlichen Datensatz an. Es handelt sich dabei um Erträge der Maisproduktion auf verschiedenen Anbauflächen in Argentinien. Sie untersuchen die Fragestellung, ob es sich lohnt, die Menge des auf die Felder gebrachten Stickstoffdüngers für jede Anbaufläche separat festzulegen. Im Ergebnis ist dies tatsächlich der Fall, auch wenn der dadurch zu erwartende zusätzliche Ertrag vergleichsweise gering ausfällt. Ezcurra et al. (2006) untersuchen die verschiedenen Regionen der europäischen Union auf ihre Wohlfahrt (gemessen als Funktion von Pro-Kopf-BIP, Bevölkerungsentwicklung und Gleichheit der Einkommensverteilung). Auch hier entsprechen die aus der generalisierten Momentenmethode gewonnenen Ergebnisse im wesentlichen den Ergebnissen bei Verwendung der Maximum-Likelihood-Methode. Druska und Horrace (2004) entwickeln den Schätzer weiter, um ihn auf Paneldaten zum Reisanbau in Indonesien im Zeitraum von 1977 bis 1983 anwenden zu können. Kelejian und Prucha (2004) verallgemeinern den von ihnen vorgeschlagenen Schätzer auf simultane Gleichungssysteme.

Der verallgemeinerte Momentenmethodenschätzer von Kelejian und Prucha (1999) wird also vielfach verwendet. Außerdem existieren Weiterentwicklungen in verschiedene Richtungen wie Paneldaten oder simultane Gleichungen. Das nächste Unterkapitel stellt eine Modifikation des Schätzers vor, welche den wesentlichen Vorteil (geringer Rechenaufwand) beibehält und die Eigenschaften des Schätzers verbessert.

3.2 Ein alternativer Momentenmethodenschätzer

Die Grundidee zur Verbesserung des verallgemeinerten Momentenmethodenschätzers aus dem letzten Unterkapitel ist die folgende: Der Schätzer setzt die theoretischen Momente der Störgrößen u mit empirischen Momenten der Residuen \hat{u} gleich. Dabei müssen die empirischen Momente der Residuen verwendet werden, weil die Störgrößen nicht beobachtbar sind. Der Unterschied zwischen Störgrößen und Residuen ist zwar asymptotisch zu vernachlässigen. Für jeden festen Stichprobenumfang jedoch sind die Residuen nicht exakt gleich den Störgrößen. Wenn für die Konstruktion des Schätzers nun die Störgrößen ohnehin durch die Residuen ersetzt werden müssen, dann sollte man von vornherein die theoretischen Momente der Residuen berechnen und nicht diejenigen der Störgrößen. Dabei wird ausgenutzt, dass der Vektor \hat{u} senkrecht auf den Spalten der Matrix X steht, denn es gilt die Beziehung $\hat{u} = Mu$ mit $M = I - X(X^T X)^{-1} X^T$. Als orthogonaler Projektor ist M idempotent und symmetrisch, so dass die theoretischen Momente der Residuen folgendermaßen lauten:

$$\mathbb{E} \left(\frac{1}{n} (M\varepsilon)^T M\varepsilon \right) = \frac{\sigma^2}{n} \text{tr}(M), \quad (9)$$

$$\mathbb{E} \left(\frac{1}{n} (WM\varepsilon)^T WM\varepsilon \right) = \frac{\sigma^2}{n} \text{tr}(MW^T W), \quad (10)$$

$$\mathbb{E} \left(\frac{1}{n} (WM\varepsilon)^T M\varepsilon \right) = \frac{\sigma^2}{n} \text{tr}(WM). \quad (11)$$

Dabei wird ausgenutzt, dass sich die Spur eines Matrixproduktes bei zyklischer Vertauschung der Matrizen nicht ändert, vgl. S. 61 in Schmidt und Trenkler (2006). Multiplikation von (1) mit M bzw. WM liefert die Beziehungen

$$M\varepsilon = Mu - \rho MWu, \quad (12)$$

$$WM\varepsilon = WMu - \rho WMWu. \quad (13)$$

Einsetzen der Gleichungen (12) und (13) in die Momentenbedingungen (9)-(11) führt zu

$$\begin{aligned} \frac{1}{n}\mathbf{E}(u^T Mu) - \frac{2\rho}{n}\mathbf{E}(u^T MWu) \\ + \frac{\rho^2}{n}\mathbf{E}(u^T W^T MWu) &= \frac{\sigma^2}{n}\text{tr}(M), \\ \frac{1}{n}\mathbf{E}(u^T MW^T WMu) - \frac{2\rho}{n}\mathbf{E}(u^T W^T WMWu) \\ + \frac{\rho^2}{n}\mathbf{E}(u^T W^T MW^T WMWu) &= \frac{\sigma^2}{n}\text{tr}(MW^T W), \\ \frac{1}{n}\mathbf{E}(u^T MW^T Mu) - \frac{\rho}{n}\mathbf{E}(u^T M[W + W^T]MWu) \\ + \frac{\rho^2}{n}\mathbf{E}(u^T W^T MW MWu) &= \frac{\sigma^2}{n}\text{tr}(WM). \end{aligned}$$

Es ergibt sich das theoretische Gleichungssystem

$$\Gamma \begin{pmatrix} \rho \\ \rho^2 \\ \sigma^2 \end{pmatrix} - \gamma = 0. \quad (14)$$

Dabei hat die (3×3) -Matrix Γ die Einträge

$$\begin{aligned}
\Gamma_{(1,1)} &= \frac{2}{n} \mathbf{E} (u^T M W u), \\
\Gamma_{(1,2)} &= -\frac{1}{n} \mathbf{E} (u^T W^T M W u), \\
\Gamma_{(1,3)} &= \frac{1}{n} \text{tr}(M), \\
\Gamma_{(2,1)} &= \frac{2}{n} \mathbf{E} (u^T W^T W M W M u), \\
\Gamma_{(2,2)} &= -\frac{1}{n} \mathbf{E} (u^T W^T M W^T W M W u), \\
\Gamma_{(2,3)} &= \frac{1}{n} \text{tr} (M W^T W), \\
\Gamma_{(3,1)} &= \frac{1}{n} \mathbf{E} (u^T M [W + W^T] M W u), \\
\Gamma_{(3,2)} &= -\frac{1}{n} \mathbf{E} (u^T W^T M W M W u), \\
\Gamma_{(3,3)} &= \frac{1}{n} \text{tr}(W M)
\end{aligned}$$

und der (3×1) -Vektor γ lautet

$$\gamma = \left(\frac{1}{n} \mathbf{E} (u^T M u), \frac{1}{n} \mathbf{E} (u^T M W^T W M u), \frac{1}{n} \mathbf{E} (u^T M W^T M u) \right)^T.$$

Schließlich werden für jede $(n \times n)$ -Matrix A die theoretischen Momente $\mathbf{E}(u^T A u)$ durch die empirischen Momente $\hat{u}^T A \hat{u}$ ersetzt. Mit $M u = \hat{u}$ und $\text{tr}(M) = n - k$ lautet das empirische Gegenstück zu (14)

$$H \begin{pmatrix} \rho \\ \rho^2 \\ \sigma^2 \end{pmatrix} - h =: w(\rho, \sigma^2),$$

wobei

$$H = \begin{pmatrix} \frac{2}{n}\hat{u}^T W \hat{u} & -\frac{1}{n}\hat{u}^T W^T M W \hat{u} & \frac{n-k}{n} \\ \frac{2}{n}\hat{u}^T W^T W M W \hat{u} & -\frac{1}{n}\hat{u}^T W^T M W^T W M W \hat{u} & \frac{1}{n}\text{tr}(M W^T W) \\ \frac{1}{n}\hat{u}^T [W + W^T] M W \hat{u} & -\frac{1}{n}\hat{u}^T W^T M W M W \hat{u} & \frac{1}{n}\text{tr}(W M) \end{pmatrix} \quad (15)$$

und

$$h = \left(\frac{1}{n}\hat{u}^T \hat{u}, \frac{1}{n}\hat{u}^T W^T W \hat{u}, \frac{1}{n}\hat{u}^T W \hat{u} \right)^T.$$

Der Vektor $w(\rho, \sigma^2)$ bildet das empirische Gegenstück zum Nullvektor in (14) und die Schätzer für ρ und σ^2 sind diejenigen Werte, welche die euklidische Länge des Vektors $w(\rho, \sigma^2)$ minimieren. Der modifizierte nichtlineare KQ-Schätzer für ρ und σ^2 nach der generalisierten Momentenmethode lautet somit formal

$$(\hat{\rho}_{NLS}^*, \hat{\sigma}_{NLS}^{2*})^T = \operatorname{argmin}\{w(\rho, \sigma^2)^T w(\rho, \sigma^2) : \rho \in [-a, a], \sigma^2 \in [0, b]\}, \quad (16)$$

wobei $a \geq 1$ und $b < \infty$. Der zulässige Bereich für σ^2 ist das abgeschlossene Intervall $[0, b]$, obwohl der Varianzparameter in der Praxis immer positiv sein wird. Die Aufnahme des Wertes 0 sorgt dafür, dass der Parameterraum kompakt ist. Die Konstanten a und b begrenzen den Parameterraum von ρ und σ^2 . In der Praxis ist bekannt, in welcher Größenordnung die wahren Werte von ρ und σ^2 liegen. Die Konstanten a und b können dann hinreichend groß gewählt werden, so dass der Parameterraum die wahren Werte mit Sicherheit enthält. Die formale Einschränkung des Parameterraums sowie seine Kompaktheit sind für den Nachweis der Konsistenz im nächsten Unterkapitel erforderlich.

3.3 Konsistenz des alternativen Momentenmethodenschätzers

Die folgenden beiden technischen Annahmen seien erfüllt:

Annahme 5 Für den kleinsten Eigenwert der Matrix $\Gamma^T \Gamma$ gilt $\lambda_{\min}(\Gamma^T \Gamma) \geq \lambda_* > 0$, wobei λ_* von ρ und σ^2 abhängen darf.

Annahme 6 Die Zeilen- und Spaltensummen der Beträge der Einträge der Matrix $M = I - X(X^T X)^{-1} X^T$ sind beschränkt, $\sum_{j=1}^n |m_{ij}| \leq c_m \forall i = 1, \dots, n$, $\sum_{i=1}^n |m_{ij}| \leq c_m \forall j = 1, \dots, n$ mit $0 < c_m < \infty$.

Annahme 5 stellt die Identifizierbarkeit von ρ und σ^2 aus den Momentenbedingungen sicher: Mit dem kleinsten Eigenwert der Matrix $\Gamma^T \Gamma$ sind auch die Eigenwerte der Matrix Γ selbst von Null verschieden, so dass die theoretischen Momentenbedingungen (14) eine eindeutige Lösung in ρ und σ^2 besitzen. Annahme 6 ermöglicht die Konvergenz der empirischen Momente der Residuen gegen die theoretischen. Theorem 3.1 stellt die Konsistenz des GMM-Schätzers dar. Der Beweis überträgt die Beweisideen aus Kelejian und Prucha (1999) auf den modifizierten Schätzer, wobei die einzelnen Schritte hier ausführlicher dargestellt werden.

Theorem 3.1 Sei $(\hat{\rho}_{NLS}^*, \hat{\sigma}_{NLS}^{2*})$ der in (16) definierte Schätzer für (ρ, σ^2) . Unter den Annahmen 1-6 gilt für $n \rightarrow \infty$:

$$(\hat{\rho}_{NLS}^*, \hat{\sigma}_{NLS}^{2*}) \xrightarrow{p} (\rho, \sigma^2).$$

Für den Beweis des Theorems werden einige Lemmata benötigt. Dazu seien die folgenden Zufallsvariablen definiert:

$$\begin{aligned}
\vartheta_1 &= \frac{1}{n} \hat{u}^T W u &= \frac{1}{n} \varepsilon^T C_1 \varepsilon, \\
\vartheta_2 &= \frac{1}{n} u^T W^T M W u &= \frac{1}{n} \varepsilon^T C_2 \varepsilon, \\
\vartheta_3 &= \frac{1}{n} u^T W^T W M W \hat{u} &= \frac{1}{n} \varepsilon^T C_3 \varepsilon, \\
\vartheta_4 &= \frac{1}{n} u^T W^T M W^T W M W u &= \frac{1}{n} \varepsilon^T C_4 \varepsilon, \\
\vartheta_5 &= \frac{1}{n} \hat{u}^T [W + W^T] M W u &= \frac{1}{n} \varepsilon^T C_5 \varepsilon, \\
\vartheta_6 &= \frac{1}{n} u^T W^T M W M W u &= \frac{1}{n} \varepsilon^T C_6 \varepsilon,
\end{aligned}$$

wobei

$$\begin{aligned}
C_1 &= P^T M W P, \\
C_2 &= P^T W^T M W P, \\
C_3 &= P^T W^T W M W M P, \\
C_4 &= P^T W^T M W^T W M W P, \\
C_5 &= P^T M [W + W^T] M W P, \\
C_6 &= P^T W^T M W M W P.
\end{aligned}$$

Die Erwartungswerte der ϑ_i bilden (bis auf konstante Vorfaktoren) die ersten beiden Spalten der Matrix Γ . Die entsprechenden in der Matrix H enthaltenen Zufallsvariablen, bei denen u jeweils durch \hat{u} ersetzt wird, lauten $\hat{\vartheta}_1$ bis $\hat{\vartheta}_6$. Die in dem Vektor h enthaltenen Zufallsvariablen sind

$$\begin{aligned}
\vartheta_7 &= \frac{1}{n} \hat{u}^T \hat{u} &= \frac{1}{n} \varepsilon^T C_7 \varepsilon, & C_7 = P^T M P, \\
\vartheta_8 &= \frac{1}{n} \hat{u}^T W^T W \hat{u} &= \frac{1}{n} \varepsilon^T C_8 \varepsilon, & C_8 = P^T M W^T W M P, \\
\vartheta_9 &= \frac{1}{n} \hat{u}^T W \hat{u} &= \frac{1}{n} \varepsilon^T C_9 \varepsilon, & C_9 = P^T M W M P.
\end{aligned}$$

Der Beweis von Theorem 3.1 baut darauf auf, dass die ϑ_h für $h = 1, \dots, 9$ in Wahrscheinlichkeit gegen ihre Erwartungswerte konvergieren. Ferner konvergieren die Differenzen zwischen den $\hat{\vartheta}_h$ und den ϑ_h für $h = 1, \dots, 6$ in Wahrscheinlichkeit gegen Null und somit auch diese $\hat{\vartheta}_h$ in Wahrscheinlichkeit gegen die Erwartungswerte der ϑ_h , so dass - ebenfalls in Wahrscheinlichkeit - H gegen Γ und h gegen γ konvergiert. Als letztes wird gezeigt, dass die Parameter ρ und σ^2 aus der Beziehung $\Gamma(\rho, \rho^2, \sigma^2)^T - \gamma = 0$ eindeutig identifizierbar sind.

Lemma 3.2 Seien A und B zwei Matrizen mit beschränkten Zeilen- und Spaltensummen der Beträge der Einträge, $\sum_{j=1}^n |a_{ij}| \leq c_a \forall i = 1, \dots, n$, $\sum_{i=1}^n |a_{ij}| \leq c_a \forall j = 1, \dots, n$, $\sum_{j=1}^n |b_{ij}| \leq c_b \forall i = 1, \dots, n$, $\sum_{i=1}^n |b_{ij}| \leq c_b \forall j = 1, \dots, n$. Dann sind auch die Zeilen- und Spaltensummen der Beträge der Einträge der Matrizen $D = A + B$ und $E = AB$ beschränkt.

Beweis: Wegen $|d_{ij}| = |a_{ij} + b_{ij}| \leq |a_{ij}| + |b_{ij}|$ sind die Zeilen- und Spaltensummen der Beträge der Einträge von D jeweils durch $c_a + c_b$ beschränkt. Für das Matrixprodukt ergibt sich

$$\sum_{j=1}^n |e_{ij}| = \sum_{j=1}^n \left| \sum_{k=1}^n a_{ik} b_{kj} \right| \leq \sum_{j=1}^n \sum_{k=1}^n |a_{ik}| |b_{kj}| = \sum_{k=1}^n |a_{ik}| \sum_{j=1}^n |b_{kj}| \leq c_a c_b$$

und für die Spaltensummen gilt entsprechend auch $\sum_{i=1}^n |e_{ij}| \leq c_a c_b$. \square

Lemma 3.3 Unter den Annahmen 1(c), 1(d), 3 und 6 haben die Einträge der Matrizen C_h für $h = 1, \dots, 9$ die folgenden Eigenschaften:

- a) $\sum_{j=1}^n |c_{h,ij}| \leq c \forall i = 1, \dots, n$, $\sum_{i=1}^n |c_{h,ij}| \leq c \forall j = 1, \dots, n$ mit $0 < c < \infty$.
- b) $\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (c_{h,ij} + c_{h,ji})^2 = o(1)$.

Beweis: Die Aussage (a) folgt aus den Annahmen 1(c), 1(d), 3 und 6 sowie Lemma 3.2. Folglich sind die Zeilen- und Spaltensummen der Beträge der Einträge der Matrix $C_h + C_h^T$ durch $2c$ beschränkt und die Summe der Beträge der Hauptdiagonaleinträge von $(C_h + C_h^T)(C_h + C_h^T)$ ist nicht größer als $4nc^2$. Da $\sum_{j=1}^n (c_{h,ij} + c_{h,ji})^2$ dem i -ten Hauptdiagonaleintrag der Matrix $(C_h + C_h^T)(C_h + C_h^T)$ entspricht, ist

$$\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (c_{h,ij} + c_{h,ji})^2 = \frac{1}{n^2} \text{tr} [(C_h + C_h^T)(C_h + C_h^T)] \leq \frac{4c^2}{n} \xrightarrow{n \rightarrow \infty} 0,$$

womit auch die zweite Aussage des Lemmas bewiesen ist. \square

Lemma 3.4 *Sei C eine reellwertige $(n \times n)$ -Matrix und ε ein n -dimensionaler Zufallsvektor mit $E(\varepsilon) = 0$, $\text{Cov}(\varepsilon) = \sigma^2 I$ und $E(\varepsilon_i^4) = \mu_4$, $i = 1, \dots, n$. Dann gilt*

$$\begin{aligned} E(\varepsilon^T C \varepsilon) &= \sigma^2 \sum_{i=1}^n c_{ii} = \sigma^2 \text{tr}(C), \\ \text{Var}(\varepsilon^T C \varepsilon) &= (\mu_4 - \sigma^4) \sum_{i=1}^n c_{ii}^2 + \sigma^4 \sum_{i=1}^{n-1} \sum_{j=i+1}^n (c_{ij} + c_{ji})^2. \end{aligned}$$

Beweis: Wegen $\varepsilon^T C \varepsilon = \varepsilon^T \left[\frac{1}{2}(C + C^T) \right] \varepsilon$ können aufgrund der Symmetrie der Matrix $\frac{1}{2}(C + C^T)$ die Formeln für die Momente quadratischer Formen in einer symmetrischen Matrix aus Theorem 1.5 auf S. 9 und Theorem 1.6 auf S. 10 in Seber und Lee (2003) angewendet werden. Danach gilt

$$E(\varepsilon^T C \varepsilon) = \text{tr} \left[\frac{1}{2}(C + C^T) \text{Cov}(\varepsilon) \right] = \frac{\sigma^2}{2} [\text{tr}(C) + \text{tr}(C^T)] = \sigma^2 \text{tr}(C).$$

Da die Hauptdiagonaleinträge der Matrizen C und $\frac{1}{2}(C + C^T)$ übereinstimmen, ergibt sich für die Varianz

$$\begin{aligned} \text{Var}(\varepsilon^T C \varepsilon) &= (\mu_4 - 3\sigma^4) \sum_{i=1}^n c_{ii}^2 + 2\sigma^4 \text{tr} \left(\left[\frac{1}{2}(C + C^T) \right]^2 \right) \\ &= (\mu_4 - 3\sigma^4) \sum_{i=1}^n c_{ii}^2 \\ &\quad + \frac{\sigma^4}{2} [\text{tr}(CC) + \text{tr}(CC^T) + \text{tr}(C^T C) + \text{tr}(C^T C^T)] \\ &= (\mu_4 - 3\sigma^4) \sum_{i=1}^n c_{ii}^2 + \sigma^4 [\text{tr}(CC) + \text{tr}(CC^T)], \end{aligned}$$

wobei ausgenutzt wird, dass $\text{tr}(CC) = \text{tr}[(CC)^T] = \text{tr}(C^T C^T)$ sowie $\text{tr}(C^T C) = \text{tr}(CC^T)$ gilt. Damit ergibt sich schließlich

$$\begin{aligned}
\text{Var}(\varepsilon^T C \varepsilon) &= (\mu_4 - 3\sigma^4) \sum_{i=1}^n c_{ii}^2 + \sigma^4 \left[\sum_{i=1}^n \sum_{j=1}^n (c_{ij}c_{ji} + c_{ij}^2) \right] \\
&= (\mu_4 - 3\sigma^4) \sum_{i=1}^n c_{ii}^2 \\
&\quad + 2\sigma^4 \sum_{i=1}^n c_{ii}^2 + \sigma^4 \sum_{i=1}^{n-1} \sum_{j=i+1}^n (c_{ij}c_{ji} + c_{ij}^2 + c_{ji}c_{ij} + c_{ji}^2) \\
&= (\mu_4 - \sigma^4) \sum_{i=1}^n c_{ii}^2 + \sigma^4 \sum_{i=1}^{n-1} \sum_{j=i+1}^n (c_{ij} + c_{ji})^2. \quad \square
\end{aligned}$$

Lemma 3.5 *Unter den Annahmen 1 - 4 sowie 6 haben die Zufallsvariablen ϑ_h , $h = 1, \dots, 9$, die folgenden Eigenschaften:*

- a) $E(\vartheta_h) = O(1)$
- b) $\text{Var}(\vartheta_h) = o(1)$
- c) $\vartheta_h - E(\vartheta_h) \xrightarrow{p} 0$ für $n \rightarrow \infty$
- d) $\vartheta_h = O_p(1)$

Beweis: Wegen Lemma 3.3(a) sind die Beträge der Einträge der Matrizen C_h allesamt beschränkt, so dass mit Lemma 3.4 für die Erwartungswerte

$$E(\vartheta_h) = \frac{\sigma^2}{n} \sum_{i=1}^n c_{h,ii} = O(1)$$

gilt. Ebenfalls mit Lemma 3.4 ergibt sich für die Varianzen

$$\text{Var}(\vartheta_h) = \frac{1}{n^2} \left[(\mu_4 - \sigma^4) \sum_{i=1}^n c_{h,ii}^2 + \sigma^4 \sum_{i=1}^{n-1} \sum_{j=i+1}^n (c_{h,ij} + c_{h,ji})^2 \right],$$

und da nach Lemma 3.3(b)

$$\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (c_{h,ij} + c_{h,ji})^2 = o(1)$$

gilt, ist $\text{Var}(\vartheta_h) = o(1)$. Teil (c) folgt mit den Aussagen (a) und (b) aus der Chebyschev'schen Ungleichung. Da nach (a) der Erwartungswert der ϑ_h beschränkt ist und nach (b) die Varianz gegen Null konvergiert, ergibt sich Aussage (d) aus Folgerung 5.1.1.2 auf Seite 220 in Fuller (1996). \square

Lemma 3.6 *Seien v_i, w_i, \hat{v}_i und \hat{w}_i Zufallsvariablen, für die*

$$|\hat{v}_i - v_i| \leq D_{v,i} \tau_v \quad \text{und} \quad |\hat{w}_i - w_i| \leq D_{w,i} \tau_w$$

gilt. Dabei sind $D_{v,i}, D_{w,i}, \tau_v$ und τ_w nicht negative Zufallsvariablen mit den Eigenschaften $\frac{1}{n} \sum_{i=1}^n D_{v,i}^2 = O_p(1), \frac{1}{n} \sum_{i=1}^n D_{w,i}^2 = O_p(1), \tau_v = o_p(1), \tau_w = o_p(1)$. Weiter gelte $\frac{1}{n} \sum_{i=1}^n v_i^2 = O_p(1)$ und $\frac{1}{n} \sum_{i=1}^n w_i^2 = O_p(1)$. Dann gilt für $n \rightarrow \infty$:

$$\frac{1}{n} \sum_{i=1}^n \hat{v}_i \hat{w}_i - \frac{1}{n} \sum_{i=1}^n v_i w_i \xrightarrow{p} 0.$$

Beweis: Wegen

$$\begin{aligned} \hat{v}_i \hat{w}_i - v_i w_i &= \hat{v}_i \hat{w}_i - v_i w_i + \hat{v}_i w_i - \hat{v}_i w_i + \hat{w}_i v_i - \hat{w}_i v_i + v_i w_i - v_i w_i \\ &= \hat{v}_i w_i - v_i w_i + \hat{w}_i v_i - v_i w_i + \hat{v}_i \hat{w}_i - \hat{v}_i w_i - \hat{w}_i v_i + v_i w_i \\ &= (\hat{v}_i - v_i) w_i + (\hat{w}_i - w_i) v_i + (\hat{v}_i - v_i)(\hat{w}_i - w_i) \end{aligned}$$

gilt aufgrund der Dreiecksungleichung

$$\begin{aligned}
\left| \frac{1}{n} \sum_{i=1}^n \hat{v}_i \hat{w}_i - \frac{1}{n} \sum_{i=1}^n v_i w_i \right| &\leq \frac{1}{n} \sum_{i=1}^n |\hat{v}_i - v_i| |w_i| + \frac{1}{n} \sum_{i=1}^n |\hat{w}_i - w_i| |v_i| \\
&\quad + \frac{1}{n} \sum_{i=1}^n |\hat{v}_i - v_i| |\hat{w}_i - w_i| \\
&\leq \frac{1}{n} \sum_{i=1}^n D_{v,i} |w_i| \tau_v + \frac{1}{n} \sum_{i=1}^n D_{w,i} |v_i| \tau_w \\
&\quad + \frac{1}{n} \sum_{i=1}^n D_{v,i} \tau_v D_{w,i} \tau_w \\
&\leq \left(\frac{1}{n} \sum_{i=1}^n D_{v,i}^2 \right)^{\frac{1}{2}} \left(\frac{1}{n} \sum_{i=1}^n w_i^2 \right)^{\frac{1}{2}} \tau_v \\
&\quad + \left(\frac{1}{n} \sum_{i=1}^n D_{w,i}^2 \right)^{\frac{1}{2}} \left(\frac{1}{n} \sum_{i=1}^n v_i^2 \right)^{\frac{1}{2}} \tau_w \\
&\quad + \left(\frac{1}{n} \sum_{i=1}^n D_{v,i}^2 \right)^{\frac{1}{2}} \left(\frac{1}{n} \sum_{i=1}^n D_{w,i}^2 \right)^{\frac{1}{2}} \tau_v \tau_w.
\end{aligned}$$

Dabei folgt das zweite Ungleichheitszeichen aus den Voraussetzungen des Lemmas und das dritte aus der Hölderschen Ungleichung (vgl. Marshall und Olkin (1979), Seite 457). Da nach Voraussetzung die Ausdrücke in den Klammern allesamt in Wahrscheinlichkeit beschränkt sind und $\tau_v = o_p(1)$ sowie $\tau_w = o_p(1)$ gilt, ist die Aussage des Lemmas bewiesen. \square

Lemma 3.7 *Unter den Annahmen 1-4 und 6 gilt für $h = 1, \dots, 6$ und $n \rightarrow \infty$*

$$\hat{\vartheta}_h - \vartheta_h \xrightarrow{p} 0.$$

Beweis: Es wird gezeigt, dass die Einträge der Vektoren \hat{u} und u , $\hat{u}^{(1)} = W\hat{u}$ und $u^{(1)} = Wu$, $\hat{u}^{(2)} = MW\hat{u}$ und $u^{(2)} = MWu$ sowie $\hat{u}^{(3)} = WMW\hat{u}$ und $u^{(3)} = WMWu$ jeweils die Bedingungen für \hat{v}_i und v_i aus Lemma 3.6 erfüllen. Nach Lemma 3.5(d) gilt

$$\frac{1}{n} \sum_{i=1}^n \left(u_i^{(2)}\right)^2 = \vartheta_2 = O_p(1)$$

sowie

$$\frac{1}{n} \sum_{i=1}^n \left(u_i^{(3)}\right)^2 = \vartheta_4 = O_p(1).$$

Mit der Argumentation aus dem Beweis von Lemma 3.5(a) folgt weiter

$$\frac{1}{n} \sum_{i=1}^n \left(u_i^{(1)}\right)^2 = O_p(1)$$

sowie

$$\frac{1}{n} \sum_{i=1}^n (u_i)^2 = O_p(1),$$

da die Matrix $W^T W$ sowie trivialerweise auch die Einheitsmatrix die in Lemma 3.3 aufgeführten Eigenschaften der Matrizen C_h aufweisen. Nach Annahme 4 gilt außerdem $|\hat{u}_i - u_i| \leq \|d_i\| \|\Delta\|$ mit $\frac{1}{n} \sum_{i=1}^n \|d_i\|^{2+\delta} = O_p(1)$ für ein $\delta > 0$ und $\sqrt{n} \|\Delta\| = O_p(1)$. Da nach der Ungleichung von Ljapunow (vgl. Marshall und Olkin (1979), Seite 459, hier angewendet mit $a = 2 + \delta$, $b = 2$ und $c = 0$)

$$\frac{1}{n} \sum_{i=1}^n \|d_i\|^2 \leq \left(\frac{1}{n} \sum_{i=1}^n \|d_i\|^{2+\delta} \right)^{\frac{2}{2+\delta}}$$

gilt, sind die Voraussetzungen von Lemma 3.6 für \hat{u} und u mit $D_{v,i} = \|d_i\|$ und $\tau_v = \|\Delta\|$ erfüllt.

Als nächstes wird gezeigt, dass Lemma 3.6 auch auf $\hat{u}^{(1)}$ und $u^{(1)}$ anwendbar ist. Betrachte dazu

$$\sum_{i=1}^n |w_{ij}|^p = c_w^{p-1} \sum_{i=1}^n |w_{ij}| \left(\frac{|w_{ij}|}{c_w} \right)^{p-1} \leq c_w^{p-1} \sum_{i=1}^n |w_{ij}| \leq c_w^p$$

sowie $\hat{u}^{(1)} - u^{(1)} = W(\hat{u} - u)$, woraus

$$\hat{u}_i^{(1)} - u_i^{(1)} = \sum_{j=1}^n w_{ij}(\hat{u}_j - u_j) \quad (17)$$

folgt. Anwendung der Dreiecksungleichung in Verbindung mit Annahme 4 sowie der Ungleichung von Hölder mit $q = 2 + \delta$ und $\frac{1}{q} + \frac{1}{p} = 1$ liefert

$$\begin{aligned} \left| \hat{u}_i^{(1)} - u_i^{(1)} \right| &\leq \sum_{j=1}^n |w_{ij}| \|d_j\| \|\Delta\| \\ &\leq \left(\sum_{j=1}^n |w_{ij}|^p \right)^{\frac{1}{p}} \left(\sum_{j=1}^n \|d_j\|^q \right)^{\frac{1}{q}} \|\Delta\| \\ &\leq c_w \left(\frac{1}{n} \sum_{j=1}^n \|d_j\|^q \right)^{\frac{1}{q}} n^{\frac{1}{q}} \|\Delta\| \\ &= D^{(1)} \tau^{(1)} \end{aligned}$$

mit

$$D^{(1)} = c_w \left(\frac{1}{n} \sum_{j=1}^n \|d_j\|^q \right)^{\frac{1}{q}}$$

und $\tau^{(1)} = n^{\frac{1}{q}} \|\Delta\|$. Nach Annahme 4 ist

$$\frac{1}{n} \sum_{j=1}^n \|d_j\|^q = O_p(1)$$

und $\sqrt{n} \|\Delta\| = O_p(1)$, so dass Lemma 3.6 wegen $q = 2 + \delta$ auch auf $\hat{u}^{(1)}$ und $u^{(1)}$ anwendbar ist.

Weiter ist $\hat{u}^{(2)} - u^{(2)} = M(\hat{u}^{(1)} - u^{(1)})$ und $\hat{u}^{(3)} - u^{(3)} = W(\hat{u}^{(2)} - u^{(2)})$, so dass sich wegen (17)

$$\begin{aligned} \hat{u}_i^{(2)} - u_i^{(2)} &= \sum_{j=1}^n m_{ij} \sum_{s=1}^n w_{js} (\hat{u}_s - u_s), \\ \hat{u}_i^{(3)} - u_i^{(3)} &= \sum_{j=1}^n w_{ij} \sum_{s=1}^n m_{js} \sum_{t=1}^n w_{st} (\hat{u}_t - u_t) \end{aligned}$$

ergibt. Damit gilt aufgrund der Annahmen 1(c), 1(d), 4 und 6

$$\begin{aligned} \left| \hat{u}_i^{(2)} - u_i^{(2)} \right| &\leq \sum_{j=1}^n |m_{ij}| \sum_{s=1}^n |w_{js}| \|d_s\| \|\Delta\| \\ &\leq c_m \left(\sum_{s=1}^n |w_{js}|^p \right)^{\frac{1}{p}} \left(\sum_{s=1}^n \|d_s\|^q \right)^{\frac{1}{q}} \|\Delta\| \\ &\leq c_m c_w \left(\frac{1}{n} \sum_{s=1}^n \|d_s\|^q \right)^{\frac{1}{q}} n^{\frac{1}{q}} \|\Delta\| \\ &= D^{(2)} \tau^{(2)} \end{aligned}$$

sowie

$$\begin{aligned}
\left| \hat{u}_i^{(3)} - u_i^{(3)} \right| &\leq \sum_{j=1}^n |w_{ij}| \sum_{s=1}^n |m_{js}| \sum_{t=1}^n |w_{st}| \|d_t\| \|\Delta\| \\
&\leq c_w c_m \left(\sum_{t=1}^n |w_{st}|^p \right)^{\frac{1}{p}} \left(\sum_{t=1}^n \|d_t\|^q \right)^{\frac{1}{q}} \|\Delta\| \\
&\leq c_w^2 c_m \left(\frac{1}{n} \sum_{t=1}^n \|d_t\|^q \right)^{\frac{1}{q}} n^{\frac{1}{q}} \|\Delta\| \\
&= D^{(3)} \tau^{(3)}.
\end{aligned}$$

Dabei erfüllen

$$D^{(2)} = c_m c_w \left(\frac{1}{n} \sum_{s=1}^n \|d_s\|^q \right)^{\frac{1}{q}},$$

$$D^{(3)} = c_w^2 c_m \left(\frac{1}{n} \sum_{t=1}^n \|d_t\|^q \right)^{\frac{1}{q}}$$

und $\tau^{(2)} = \tau^{(3)} = n^{\frac{1}{q}} \|\Delta\|$ genau wie $D^{(1)}$ und $\tau^{(1)}$ die Voraussetzungen von Lemma 3.6, so dass das Lemma bewiesen ist. \square

Beweis von Theorem 3.1. Die zu minimierende reellwertige Funktion

$$w(\rho, \sigma^2)^T w(\rho, \sigma^2)$$

ist für festes $(\rho, \sigma^2)^T$ eine messbare Funktion von u und für festes u stetig in ρ und σ^2 . Außerdem ist der Parameterraum $[-a, a] \times [0, b]$ kompakt. Dies sind die beiden Voraussetzungen von Lemma 2 in Jennrich (1969), so dass

das Lemma anwendbar und somit zunächst einmal Existenz und Messbarkeit von $\hat{\rho}_{NLS}^*$ und $\hat{\sigma}_{NLS}^{2*}$ gewährleistet sind.

Die Zielfunktion $R(\theta)$ des nichtlinearen KQ-Schätzers $(\hat{\rho}_{NLS}^*, \hat{\sigma}_{NLS}^{2*})^T$ sowie das nichtstochastische Gegenstück $\bar{R}(\theta)$ seien definiert als

$$\begin{aligned} R(\underline{\theta}) &= [H(\underline{\rho}, \underline{\rho}^2, \underline{\sigma}^2)^T - h]^T [H(\underline{\rho}, \underline{\rho}^2, \underline{\sigma}^2)^T - h], \\ \bar{R}(\underline{\theta}) &= [\Gamma(\underline{\rho}, \underline{\rho}^2, \underline{\sigma}^2)^T - \gamma]^T [\Gamma(\underline{\rho}, \underline{\rho}^2, \underline{\sigma}^2)^T - \gamma]. \end{aligned}$$

Dabei ist $\underline{\theta} = (\underline{\rho}, \underline{\sigma}^2)^T$. Zum Beweis des Theorems wird gezeigt, dass die Voraussetzungen von Lemma 3.1 in Pötscher und Prucha (1991) erfüllt sind. Diese sind:

- a) Der wahre Parametervektor $\theta = (\rho, \sigma^2)$ ist eindeutig identifizierbar, d.h., der wahre Parametervektor θ minimiert die (nichtstochastische) Zielfunktion \bar{R} bezüglich $\underline{\theta}$.
- b) Der Schätzer ist definiert als Minimum der (aus den Daten bestimmten) Zielfunktion $R(\underline{\theta})$.
- c) Der maximale Abstand zwischen der wahren Zielfunktion \bar{R} und der aus den Daten gewonnenen Zielfunktion R konvergiert gegen Null.

Zum Nachweis von (a) wird zunächst verwendet, dass aus Gleichung (14)

$$\gamma = \Gamma(\rho, \rho^2, \sigma^2)^T$$

und somit $\bar{R}(\theta) = 0$ folgt, so dass

$$\begin{aligned} \left[\Gamma(\underline{\rho}, \underline{\rho}^2, \underline{\sigma}^2)^T - \gamma \right] &= \left[\Gamma(\underline{\rho}, \underline{\rho}^2, \underline{\sigma}^2)^T - \Gamma(\rho, \rho^2, \sigma^2)^T \right] \\ &= \left[\Gamma(\underline{\rho} - \rho, \underline{\rho}^2 - \rho^2, \underline{\sigma}^2 - \sigma^2)^T \right] \end{aligned}$$

und somit

$$\begin{aligned}
\bar{R}(\underline{\theta}) - \bar{R}(\theta) &= \left[\Gamma (\underline{\rho} - \rho, \underline{\rho}^2 - \rho^2, \underline{\sigma}^2 - \sigma^2)^T \right]^T \left[\Gamma (\underline{\rho} - \rho, \underline{\rho}^2 - \rho^2, \underline{\sigma}^2 - \sigma^2)^T \right] \\
&= (\underline{\rho} - \rho, \underline{\rho}^2 - \rho^2, \underline{\sigma}^2 - \sigma^2) \Gamma^T \Gamma (\underline{\rho} - \rho, \underline{\rho}^2 - \rho^2, \underline{\sigma}^2 - \sigma^2)^T \\
&\geq \lambda_{\min} (\Gamma^T \Gamma) (\underline{\rho} - \rho, \underline{\rho}^2 - \rho^2, \underline{\sigma}^2 - \sigma^2) (\underline{\rho} - \rho, \underline{\rho}^2 - \rho^2, \underline{\sigma}^2 - \sigma^2)^T \\
&\geq \lambda_* (\underline{\rho} - \rho, \underline{\sigma}^2 - \sigma^2) (\underline{\rho} - \rho, \underline{\sigma}^2 - \sigma^2)^T \\
&= \lambda_* \|\underline{\theta} - \theta\|^2
\end{aligned}$$

gilt. Dabei folgt das erste Ungleichheitszeichen daraus, dass eine quadratische Form in einer symmetrischen Matrix immer mindestens so groß ist wie das Skalarprodukt des Vektors mit sich selbst, multipliziert mit dem kleinsten Eigenwert der Matrix. Annahme 5 gewährleistet das zweite Ungleichheitszeichen.

Damit ist die Differenz der Funktionswerte von \bar{R} an der Stelle des wahren Parametervektors θ einerseits und einer beliebigen anderen Stelle $\underline{\theta}$ andererseits zumindest proportional zur quadrierten Norm $\|\underline{\theta} - \theta\|$. Für jedes $\varepsilon > 0$ gilt also

$$\inf_{\{\underline{\theta}: \|\underline{\theta} - \theta\| \geq \varepsilon\}} [\bar{R}(\underline{\theta}) - \bar{R}(\theta)] \geq \inf_{\{\underline{\theta}: \|\underline{\theta} - \theta\| \geq \varepsilon\}} \lambda_* \|\underline{\theta} - \theta\|^2 = \lambda_* \varepsilon^2 > 0,$$

so dass θ eindeutig identifizierbar und Voraussetzung (a) erfüllt ist. Voraussetzung (b) gilt aufgrund der Definition des Schätzers in (16). Zu zeigen bleibt (c), also dass die tatsächlich verwendete Zielfunktion $R(\underline{\theta})$ auf dem gesamten Parameterraum gleichmäßig gegen die theoretische Zielfunktion $\bar{R}(\underline{\theta})$ konvergiert. Definiere dazu die partitionierten (4×3) -Matrizen $F = (H, -h)$ sowie $\Phi = (\Gamma, -\gamma)$. Damit ist

$$\begin{aligned}
H(\underline{\rho}, \underline{\rho}^2, \underline{\sigma}^2)^T - h &= F^T(\underline{\rho}, \underline{\rho}^2, \underline{\sigma}^2, 1)^T, \\
\Gamma(\underline{\rho}, \underline{\rho}^2, \underline{\sigma}^2)^T - \gamma &= \Phi^T(\underline{\rho}, \underline{\rho}^2, \underline{\sigma}^2, 1)^T
\end{aligned}$$

und somit

$$\begin{aligned}
R(\underline{\theta}) - \bar{R}(\underline{\theta}) &= (\underline{\rho}, \underline{\rho}^2, \underline{\sigma}^2, 1) F F^T (\underline{\rho}, \underline{\rho}^2, \underline{\sigma}^2, 1)^T \\
&\quad - (\underline{\rho}, \underline{\rho}^2, \underline{\sigma}^2, 1) \Phi \Phi^T (\underline{\rho}, \underline{\rho}^2, \underline{\sigma}^2, 1)^T \\
&= (\underline{\rho}, \underline{\rho}^2, \underline{\sigma}^2, 1) (F F^T - \Phi \Phi^T) (\underline{\rho}, \underline{\rho}^2, \underline{\sigma}^2, 1)^T,
\end{aligned}$$

so dass für $\rho \in [-a, a]$ und $\sigma^2 \in [0, b]$

$$\begin{aligned}
|R(\underline{\theta}) - \bar{R}(\underline{\theta})| &\leq \|(\underline{\rho}, \underline{\rho}^2, \underline{\sigma}^2, 1)\| \|F F^T - \Phi \Phi^T\| \|(\underline{\rho}, \underline{\rho}^2, \underline{\sigma}^2, 1)^T\| \\
&= \|F F^T - \Phi \Phi^T\| \|(\underline{\rho}, \underline{\rho}^2, \underline{\sigma}^2, 1)\|^2 \\
&\leq \|F F^T - \Phi \Phi^T\| (a^2 + a^4 + b^2 + 1)
\end{aligned}$$

gilt. Nach Lemma 3.5(d) sind die Einträge der Matrix F $O_P(1)$; nach Teil (a) dieses Lemmas sind die Einträge der Matrix Φ $O(1)$. Gemäß Lemma 3.7 gilt $F - \Phi \xrightarrow{p} 0$. Da die Abbildung von F auf $F F^T$ stetig ist, folgt aus dem Theorem von Slutsky (vgl. Satz 2.27 in White (1984)) auch $\|F F^T - \Phi \Phi^T\| \xrightarrow{p} 0$, so dass $R(\underline{\theta}) - \bar{R}(\underline{\theta})$ tatsächlich für $n \rightarrow \infty$ gleichmäßig gegen Null konvergiert:

$$\sup_{\rho \in [-a, a], \sigma^2 \in [0, b]} |R(\underline{\theta}) - \bar{R}(\underline{\theta})| \leq \|F F^T - \Phi \Phi^T\| (a^2 + a^4 + b^2 + 1) \xrightarrow{p} 0.$$

Da die Voraussetzungen von Lemma 3.1 in Pötscher und Prucha (1991) somit erfüllt sind, ist das Theorem bewiesen. \square

4 Parameterschätzung nach Maximum-Likelihood-Methode

Gebräuchlicher als die im letzten Kapitel behandelte GMM-Schätzung von ρ und σ^2 ist die Maximum-Likelihood-Schätzung der Parameter. Nach Gleichung (2) ist

$$\text{Cov}(y) = \text{Cov}(u) = \sigma^2(I - \rho W)^{-1}(I - \rho W^T)^{-1} = \sigma^2 P P^T$$

sowie $E(y) = X\beta$. Bei unterstellter multivariater Normalverteilung des Störgrößenvektors ist also die Likelihoodfunktion

$$L(\rho, \sigma^2, \beta) = (2\pi)^{-\frac{n}{2}} [\det(\sigma^2 P P^T)]^{-\frac{1}{2}} \cdot \exp\left(-\frac{1}{2}(y - X\beta)^T (\sigma^2 P P^T)^{-1} (y - X\beta)\right)$$

zu maximieren. Zur konkreten Berechnung der Maximum-Likelihood-Schätzer für ρ , σ^2 und β müssen numerische Optimierungsverfahren verwendet werden. Wenn der Stichprobenumfang n groß ist oder der Parametervektor β viele Komponenten hat, nimmt diese Berechnung viel Zeit in Anspruch und kann numerisch instabil werden. In diesen Fällen bietet sich eine zweistufige Schätzprozedur an. Dabei wird zunächst β geschätzt. Anschließend müsste zur Schätzung von ρ und σ^2 die Rumpflielihoodfunktion

$$L(\rho, \sigma^2) = (2\pi)^{-\frac{n}{2}} [\det(\sigma^2 P P^T)]^{-\frac{1}{2}} \exp\left(-\frac{1}{2}u^T (\sigma^2 P P^T)^{-1} u\right) \quad (18)$$

maximiert werden. Da der Störgrößenvektor $u = y - X\beta$ jedoch nicht beobachtbar ist, wird er durch den Vektor der Regressionsresiduen $\hat{u} = Mu$ ersetzt. Wie bereits im letzten Kapitel diskutiert, verschwindet der Unterschied zwischen u und \hat{u} , wenn der Stichprobenumfang immer weiter wächst.

Nichtsdestotrotz unterscheiden sich die Verteilungen von u bzw. \hat{u} für endliche Stichprobenumfänge: Während der Träger der Verteilung von u durch \mathbb{R}^n gegeben ist, wird der Vektor der Residuen so bestimmt, dass er senkrecht zu allen Spalten der Matrix X ist. Dadurch folgt \hat{u} einer singulären multivariaten Normalverteilung. Der Träger von \hat{u} ist ein $(n - k)$ -dimensionaler Unterraum des \mathbb{R}^n . Dadurch lässt sich die Likelihoodfunktion unter Verwendung der Residuen nicht ohne weiteres wie in (18) formulieren.

Dieses Problem kann jedoch behoben werden, indem man \hat{u} geeignet transformiert. Dabei soll der $(n - k)$ -dimensionale Unterraum des \mathbb{R}^n dergestalt rotiert werden, dass k Komponenten des transformierten Zufallsvektors deterministisch und die anderen $n - k$ Komponenten stochastisch sind. Eine solche Transformation gelingt mit der Spektralzerlegung von M , $M = UDU^{-1}$, wobei D eine Diagonalmatrix mit den Eigenwerten von M auf der Hauptdiagonalen und U die Matrix der zugehörigen normierten und orthogonalen Eigenvektoren ist. Multipliziert man \hat{u} von links mit U^{-1} , dann erhält man

$$E(U^{-1}\hat{u}) = 0, \quad \text{Cov}(U^{-1}\hat{u}) = \sigma^2 DU^{-1}PP^TUD.$$

Da die Matrix M ein orthogonaler Projektor ist und X vollen Spaltenrang hat, besitzt M die Eigenwerte 0 (mit Vielfachheit k) und 1 (mit Vielfachheit $n - k$). Sortiert man die Eigenwerte von M auf der Hauptdiagonalen von D der Größe nach, so sind folglich die ersten $n - k$ Hauptdiagonaleinträge gleich Eins und die letzten k Einträge gleich Null. Dadurch nehmen die letzten k Komponenten des Zufallsvektors $U^{-1}\hat{u}$ unabhängig von der konkreten Realisation von u stets den Wert Null an und enthalten somit keinerlei Information über ρ und σ^2 . Die Parameter werden stattdessen aus dem $(n - k)$ -dimensionalen Zufallsvektor

$$\tilde{u} := (U^{-1}\hat{u})_{(1,\dots,n-k)},$$

geschätzt, dessen $(n - k) \times (n - k)$ -dimensionale Kovarianzmatrix gleich

$$\text{Cov}(\tilde{u}) = \sigma^2 (DU^{-1}PP^TUD)_{(1,\dots,n-k;1,\dots,n-k)} =: \tilde{V}$$

ist. Die zu maximierende Likelihoodfunktion lautet folglich

$$\tilde{L}(\rho, \sigma^2) = (2\pi)^{-\frac{n-k}{2}} \left[\det(\tilde{V}) \right]^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\tilde{u}^T\tilde{V}^{-1}\tilde{u}\right),$$

wobei ρ und σ^2 in \tilde{V} eingehen. Der resultierende Maximum-Likelihood-Schätzer schließlich ist definiert als

$$(\hat{\rho}_{ML}^*, \hat{\sigma}_{ML}^{2*})^T = \max_{\{\rho, \sigma^2\}} \tilde{L}(\rho, \sigma^2). \quad (19)$$

5 Erweiterungen des Modells

Die Kapitel 3 und 4 behandeln Schätzer für ρ und σ^2 , welche den Unterschied zwischen nicht beobachtbaren Störgrößen und beobachtbaren Regressionsresiduen berücksichtigen. Die Störgrößen sind räumlich korreliert und die Residuen stammen aus einer gewöhnlichen Kleinst-Quadrate-Regression. Die zugrunde liegende Idee ist jedoch nicht nur in diesem speziellen Modell anwendbar. Es existieren vielmehr natürliche Erweiterungen in verschiedene Richtungen, die in diesem Kapitel vorgestellt werden.

5.1 Residuen aus verallgemeinerter Kleinst-Quadrate-Regression

Die verwendeten Residuen müssen nicht aus einer gewöhnlichen KQ-Regression stammen, sondern können auch aus einer verallgemeinerten KQ-Regression berechnet werden. Der gewöhnliche KQ-Schätzer

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

wird dabei durch den verallgemeinerten KQ-Schätzer

$$\tilde{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} y$$

ersetzt. Dieser Schätzer ist effizient, falls $\text{Cov}(u) = \sigma^2 V$ gilt.

Auf den ersten Blick mag es seltsam erscheinen, den räumlichen Korrelationsparameter ρ aus den Residuen einer verallgemeinerten KQ-Regression zu berechnen. Schließlich setzt die verallgemeinerte KQ-Regression die Kenntnis der Störgrößen-Kovarianzmatrix, $\text{Cov}(u) = \sigma^2 V$, voraus: Wenn ρ geschätzt werden muss, kann V nicht bekannt sein. Es gibt aber durchaus realistische

Situationen, in denen diese Vorgehensweise sinnvoll sein kann, wie die folgenden zwei Beispiele zeigen.

Das erste Beispiel ist eine iterative Kleinst-Quadrate-Schätzung. Dabei wird im ersten Schritt eine gewöhnliche KQ-Regression berechnet, d.h., hier ist $V = I$. Die Residuen dieser Regression liefern Schätzungen für ρ und σ^2 und somit eine Schätzung für die Kovarianzmatrix der Störgrößen. Im zweiten Schritt wird diese geschätzte Kovarianzmatrix für V eingesetzt und eine verallgemeinerte KQ-Regression durchgeführt. Die Residuen des zweiten Schrittes liefern neue Schätzungen für ρ , σ^2 und V , die als Grundlage für die verallgemeinerte KQ-Regression in Schritt 3 dienen. Dieses Prozedur wird solange wiederholt, bis sich die Schätzungen für β , ρ und σ^2 von einem Iterationsschritt zum nächsten nicht mehr wesentlich ändern. Das Verfahren konvergiert, sofern die Voraussetzungen von Lemma 1 in Oberhofer und Kmenta (1974) erfüllt sind. Die Autoren verwenden das Lemma, um die Konvergenz der oben beschriebenen iterativen Prozedur zur Bestimmung der Maximum-Likelihood-Schätzungen in Regressionsmodellen mit bestimmten Störgrößenstrukturen zu zeigen. Als Anwendungsbeispiel behandeln sie AR[1]-Störgrößen. Das Lemma ist jedoch ebenso anwendbar auf die Maximum-Likelihood-Schätzung bei räumlich korrelierten Störgrößen und auch auf die oben beschriebene Prozedur, bei der ρ und σ^2 nach generalisierter Momentenmethode geschätzt werden. Sie kann nämlich als Maximierung der Funktion

$$f(\rho, \sigma^2, \beta) = f_1(\rho, \sigma^2, \beta) + f_2(\rho, \sigma^2, \beta)$$

aufgefasst werden. Dabei bezeichnet

$$f_1(\rho, \sigma^2, \beta) = - \left[H(\rho, \rho^2, \sigma^2)^T - h \right]^T \left[H(\rho, \rho^2, \sigma^2)^T - h \right]$$

bei gegebenem β die Maximierungsaufgabe zur Schätzung von ρ und σ^2 nach

generalisierter Momentenmethode (vgl. (16)). Die Maximierung von

$$f_2(\rho, \sigma^2, \beta) = - [(X^T V^{-1} X)^{-1} X^T V^{-1} y - \beta]^T [(X^T V^{-1} X)^{-1} X^T V^{-1} y - \beta]$$

entspricht der Berechnung des verallgemeinerten KQ-Schätzers für β bei gegebenen ρ und σ^2 . Dabei gehen ρ und σ^2 in f_2 über V ein. Der Einfluss von β auf f_1 erfolgt über den Vektor der Regressionsresiduen \hat{u} , aus dem die Einträge von H und h bestimmt werden. Die Voraussetzungen von Lemma 1 aus Oberhofer und Kmenta (1974) sind somit erfüllt: Erstens existiert eine reelle Zahl s , so dass die Menge

$$S = \{(\rho, \sigma^2, \beta^T)^T \mid f(\rho, \sigma^2, \beta^T) \geq s\}$$

nichtleer und beschränkt ist. Das ist etwa für $s = f(\rho = 0, \sigma^2 = 1, \beta^T = 0^T)$ der Fall, weil f_2 nicht nach unten beschränkt ist. Zweitens ist f auf der Menge S stetig. Drittens ist der Parameterraum für $\rho \in [-a, a]$ und $\sigma^2 \in [0, b]$ beschränkt und für $\beta \in \mathbb{R}^k$ abgeschlossen, so dass S kompakt ist. Das Lemma ist somit anwendbar und das Verfahren konvergiert.

Bei dieser Vorgehensweise wird ab dem zweiten Schritt eine bestimmte Störgrößenkovarianzmatrix unterstellt, die dann anschließend aus den Residuen dieser Regression aktualisiert wird. Die gleiche Situation (VKQ-Regression mit unterstelltem ρ und anschließende Schätzung von ρ aus den Residuen) tritt auch auf, wenn Vorabinformationen über die Parameter vorliegen, etwa aus einer Voruntersuchung oder älteren Studien.

Als zweites Beispiel ist eine Situation vorstellbar, in der zwar keine Vorabinformationen über ρ vorliegen und deshalb - ausgehend von $\rho = 0$ - die Einträge auf der Nebendiagonalen von V gleich Null gewählt werden. Trotzdem kann eine verallgemeinerte KQ-Regression sinnvoll sein, wenn es Hinweise auf heteroskedastische Störgrößen gibt und die Art der Heteroskedastie bekannt ist. Das ist etwa dann der Fall, wenn es sich bei den im Vektor y zusam-

mengefassten Beobachtungen um Mittelwerte handelt, die aus verschiedenen großen Gruppen gebildet worden sind. Ein konkretes typisches Beispiel sind Arbeitslosenquoten in verschiedenen demographischen Gruppen, wobei die verschiedenen Gruppen unterschiedlich stark besetzt sind.

Die in den Kapiteln 3 und 4 vorgestellten Schätzer lassen sich leicht auf den Fall einer verallgemeinerten KQ-Regression übertragen. Die Regressionsresiduen sind

$$\begin{aligned}\tilde{u} &= y - \tilde{y} = \left[I - X (X^T V^{-1} X)^{-1} X^T V^{-1} \right] (X\beta + u) \\ &= \left[I - X (X^T V^{-1} X)^{-1} X^T V^{-1} \right] u =: \tilde{M}u,\end{aligned}$$

so dass die Matrix \tilde{M} die Rolle der Matrix M aus den Kapiteln 3 und 4 übernimmt. Ansonsten kann die Vorgehensweise grundsätzlich übernommen werden - es sei denn, die Motivation für die Anwendung des verallgemeinerten KQ-Schätzers liegt darin begründet, dass der Vektor der Innovationen ε keine skalare Kovarianzmatrix besitzt. Ein typisches Beispiel dafür sind heteroskedastische Innovationen, d.h., in dem Modell (1) ist $\text{Cov}(\varepsilon) = \sigma^2 V$ mit V als Diagonalmatrix mit möglicherweise unterschiedlichen Einträgen auf der Hauptdiagonalen. In diesem Fall lauten die Momentenbedingungen (9)-(11) für den GMM-Schätzer

$$\text{E} \left(\frac{1}{n} (M\varepsilon)^T M\varepsilon \right) = \frac{\sigma^2}{n} \text{tr}(VM), \quad (20)$$

$$\text{E} \left(\frac{1}{n} (WM\varepsilon)^T WM\varepsilon \right) = \frac{\sigma^2}{n} \text{tr}(VMW^T WM), \quad (21)$$

$$\text{E} \left(\frac{1}{n} (WM\varepsilon)^T M\varepsilon \right) = \frac{\sigma^2}{n} \text{tr}(VMW^T M). \quad (22)$$

Zur Berechnung der Momente wird ausgenutzt, dass der Zufallsvektor ε mit

der Kovarianzmatrix $\sigma^2 V$ darstellbar ist als $V^{\frac{1}{2}}\tilde{\varepsilon}$, wobei $V^{\frac{1}{2}}V^{\frac{1}{2}} = V$ und $\text{Cov}(\tilde{\varepsilon}) = \sigma^2 I$ gilt.

Durch die veränderten Momentenbedingungen ändert sich die dritte Spalte der Matrix H aus Kapitel 3; sie lautet nunmehr

$$\left(\frac{1}{n}\text{tr}(VM), \frac{1}{n}\text{tr}(VMW^TWM), \frac{1}{n}\text{tr}(VMW^TM) \right)^T.$$

Die ersten beiden Spalten von H sowie der Vektor h ändern sich nicht, sie können aus Kapitel 3 übernommen werden.

5.2 Nichtlineare Regression

Eine zweite Erweiterung des Modells ergibt sich, falls die Regressionsbeziehung zwischen y und den Spalten von X nichtlinear ist. In diesem Fall lautet das nichtlineare Regressionsmodell

$$y_i = f(x_i, \beta) + u_i.$$

Die Beobachtung y_i wird erklärt durch eine Funktion f , die von den Ausprägungen der k Kovariablen (zusammengefasst im Vektor x_i) sowie dem k -dimensionalen Parametervektor β abhängt; genau wie im linearen Fall überlagert der Störterm u_i diesen Zusammenhang. In Kompaktschreibweise lautet das Modell

$$y = F(X, \beta) + u.$$

Dabei enthält y die n Beobachtungen der abhängigen Variablen, die $(n \times k)$ -Matrix $X = (x_1, \dots, x_n)^T$ enthält die Ausprägungen der unabhängigen Variablen und die unbeobachtbaren Störgrößen u_i sind in dem $(n \times 1)$ -Vektor

u zusammengefasst. Die Funktion F bildet die Matrix X und den Parametervektor β in den \mathbb{R}^n ab, wobei die einzelnen Komponenten des Bildvektors jeweils durch $f(x_i, \beta)$ gegeben sind.

Die Funktion F sei für alle $\beta \in \mathbb{R}^k$ stetig differenzierbar nach β und X_β sei die Funktionalmatrix von F in β :

$$X_\beta = \begin{pmatrix} \frac{\partial f(x_1, \beta)}{\partial \beta_1} & \cdots & \frac{\partial f(x_1, \beta)}{\partial \beta_k} \\ \vdots & & \vdots \\ \frac{\partial f(x_n, \beta)}{\partial \beta_1} & \cdots & \frac{\partial f(x_n, \beta)}{\partial \beta_k} \end{pmatrix}.$$

Die Matrix X_β hängt von β ab. Das lineare Regressionsmodell ist in dieser Modellklasse als Spezialfall mit $F(X, \beta) = X\beta$ enthalten. Die Funktionalmatrix X_β ist dann gleich X und somit unabhängig von β .

Die Definition des Kleinst-Quadrate-Schätzers $\hat{\beta}_{NLS}$ ist die gleiche wie im linearen Regressionsmodell. Im linearen Fall wird der Abstand des Beobachtungsvektors y zum Spaltenraum der Matrix X minimiert, im nichtlinearen Fall der Abstand zwischen y und $F(X, \beta)$. Die konkrete Berechnung des Schätzers ist komplizierter als im linearen Fall, weil der Bildraum der Funktion $F(X, \beta)$ kein k -dimensionaler Unterraum, sondern allgemeiner eine k -dimensionale Mannigfaltigkeit im \mathbb{R}^n ist. Dadurch besitzt $\hat{\beta}_{NLS}$ im Gegensatz zum linearen Fall keine geschlossene Darstellung und muss durch numerische Optimierung gefunden werden.

Die Überlegungen aus den Kapiteln 3 und 4 übertragen sich auf den Fall einer nichtlinearen Regression, weil sich genau wie im linearen Fall eine wichtige Eigenschaft der Regressionsresiduen ausnutzen lässt: Der Vektor \hat{u} liegt in einem $(n - k)$ -dimensionalen Unterraum des \mathbb{R}^n , und dieser Unterraum ist bekannt. Im linearen Regressionsmodell handelt es sich um den Raum, welcher senkrecht auf den Spalten der Matrix X steht. Im nichtlinearen Fall steht \hat{u} senkrecht auf der Tangentialebene an den Bildraum von $F(X, \beta)$ im Punkt $F(X, \hat{\beta}_{NLS})$. Diese Tangentialebene entspricht dem Spaltenraum der Matrix $X_{\hat{\beta}_{NLS}}$. Der Unterraum, in dem \hat{u} liegt, hängt von $\hat{\beta}_{NLS}$ ab und ist

somit zufällig. Nachdem $\hat{\beta}_{NLS}$ berechnet wurde, ist dieser Raum jedoch genauso bekannt wie im linearen Fall der Raum senkrecht zum Spaltenraum von X . Die Information über den Raum ist genauso wertvoll wie im linearen Regressionsmodell. Die Übertragung der in den Kapiteln 3 und 4 vorgestellten Idee auf den Fall einer nichtlinearen Regression liegt somit auf der Hand: Die Matrix $X_{\hat{\beta}_{NLS}}$ ersetzt die Matrix X und

$$M_{\hat{\beta}_{NLS}} = I - X_{\hat{\beta}_{NLS}} \left(X_{\hat{\beta}_{NLS}}^T X_{\hat{\beta}_{NLS}} \right)^{-1} X_{\hat{\beta}_{NLS}}^T$$

tritt an die Stelle von M . Ansonsten entspricht die Vorgehensweise derjenigen bei einer linearen Regression.

5.3 Andere Störgrößenstrukturen

Eine dritte Erweiterung bezieht sich nur auf den ML-Schätzer. In Kapitel 4 wurde dieser für räumlich korrelierte Störgrößen vorgestellt. Das Verfahren ist jedoch auch auf beliebige andere Kovarianzstrukturen der Störgrößen anwendbar. Dazu muss lediglich die Annahme getroffen werden, dass $\text{Cov}(u) = \sigma^2 V(\rho)$ gilt, d.h., dass die Kovarianzmatrix der Störgrößen bis auf einen oder auch mehrere zu schätzende Parameter bekannt ist. Dann kann die in (18) verwendete Kovarianzmatrix räumlich korrelierter Störgrößen, $\sigma^2 P P^T$, durch $\sigma^2 V(\rho)$ ersetzt werden. Die Schätzung der Parameter aus den Residuen erfolgt genau wie bei den räumlich korrelierten Störgrößen. Damit überträgt sich die in Kapitel 4 vorgestellte Prozedur auf beliebige Störgrößenstrukturen, etwa auf AR[1]-Residuen mit Kovarianzmatrix

$$\text{Cov}(u) = \sigma^2 V = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \ddots & \ddots & \vdots \\ \rho^2 & \ddots & \ddots & \ddots & \rho^2 \\ \vdots & \ddots & \ddots & 1 & \rho \\ \rho^{n-1} & \dots & \rho^2 & \rho & 1 \end{pmatrix}$$

oder auf äquikorrelierte Residuen mit Kovarianzmatrix

$$\text{Cov}(u) = \sigma^2 V = \sigma^2 \begin{pmatrix} 1 & \rho & \dots & \dots & \rho \\ \rho & 1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 1 & \rho \\ \rho & \dots & \dots & \rho & 1 \end{pmatrix}.$$

Diese Erweiterung des Modells funktioniert für den ML-Schätzer, nicht jedoch für den GMM-Schätzer aus Kapitel 3. Das liegt daran, dass die verwendeten verallgemeinerten Momente auf der Beziehung $u = \rho W u + \varepsilon$ basieren. Diese Beziehung gilt nur für räumlich korrelierte Störgrößen, nicht jedoch für andere Störgrößenstrukturen, so dass der in Kapitel 3 vorgestellte GMM-Schätzer bei anderen Störgrößenstrukturen natürlich nicht zur Schätzung der Parameter verwendet werden kann.

Ansonsten lassen sich die in diesem Kapitel diskutierten Erweiterungen auch miteinander kombinieren. So wäre etwa eine ML-Schätzung von ρ und σ^2 für AR[1]-Störgrößen aus den Residuen einer nichtlinearen Regression denkbar, bei der die Heteroskedastie der Störgrößen in der Regression berücksichtigt wird.

6 Simulation

Die in Kapitel 3 vorgestellte Modifikation des verallgemeinerten Momentenmethodenschätzers von Kelejian und Prucha (1999) führt auf einen Schätzer, der ebenso konsistent ist wie der ursprüngliche Schätzer. Für gegen unendlich strebende Stichprobenumfänge liefern also beide Schätzer Schätzwerte, die den unbekanntem Parametern beliebig nahe kommen. Insbesondere unterscheiden sich die Resultate der beiden Schätzer dann nicht mehr wesentlich voneinander. Dies sagt jedoch noch nichts konkretes über das Verhalten der Schätzer bei gegebener (endlicher) Stichprobe aus. Dieses Kapitel vergleicht die finiten Eigenschaften der beiden Schätzer anhand einer Simulation.

Die Eigenschaften der beiden Schätzer für feste Stichprobenumfänge hängen von der Wahl der Verwandtschaftsmatrix W sowie von dem unterstellten linearen Modell, also der Modellmatrix X , ab. Da eine Simulation für alle denkbaren Matrizen W und X nicht realisierbar ist, stellt die nachfolgend beschriebene Simulation Ergebnisse für eine feste Verwandtschaftsmatrix W und eine spezielle Modellmatrix X vor, bei der der räumliche Autokorrelationsparameter ρ verschiedene Werte durchläuft.

Konkret ist die Verwandtschaftsmatrix W so konstruiert, dass jede Komponente des Störgrößenvektors mit den drei vorangehenden sowie den drei darauffolgenden Komponenten direkt zusammenhängt. Für die ersten sowie die letzten drei Komponenten wird eine Zirkelstruktur unterstellt, so dass z.B. u_1 direkt zusammenhängt mit $u_2, u_3, u_4, u_{n-2}, u_{n-1}$ und u_n . Dadurch sind in jeder Zeile von W genau 6 Einträge von Null verschieden. Um die Zeilensummen auf Eins zu normieren, sind die Einträge für zwei direkt zusammenhängenden Komponenten gleich $\frac{1}{6}$ und die Verwandtschaftsmatrix hat die folgende Gestalt:

$$W = \begin{pmatrix} 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & \dots & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} & \ddots & \ddots & & \ddots & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} & \ddots & \ddots & \ddots & & \ddots & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & \ddots & \ddots & \ddots & \ddots & & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & & \ddots & \ddots & \ddots & \ddots & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \ddots & & \ddots & \ddots & \ddots & \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \ddots & & \ddots & \ddots & \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & \dots & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 \end{pmatrix}.$$

Die Modellmatrix X entspricht der Regression auf eine Konstante sowie zwei binäre Faktoren:

$$X = \begin{pmatrix} 1 & 1 & 1 \\ \vdots & \vdots & 0 \\ \vdots & \vdots & \vdots \\ \vdots & 1 & \vdots \\ \vdots & 0 & \vdots \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

Sofern für n Vielfache von 4 gewählt werden, entspricht die Matrix X jeweils einem balancierten Versuchsplan. Die Struktur von X kann somit beibehalten werden.

Für $n = 20, 60, 100$ und 400 , $\sigma^2 = 1$ sowie Werte für ρ zwischen $-0,8$ und $0,8$ im Abstand $0,2$ werden jeweils $l = 10000$ Realisationen des Störgrößenvektors aus dem Modell (1) erzeugt, die wiederum Realisationen des Momentenmethodenschätzers $(\hat{\rho}_{NLS}, \hat{\sigma}_{NLS}^2)$ aus Kelejian und Prucha (1999) sowie der Modifikation $(\hat{\rho}_{NLS}^*, \hat{\sigma}_{NLS}^{2*})$ aus Kapitel 3 liefern. Der Vergleich der beiden Schätzer anhand der simulierten Realisationen erfolgt in zweierlei Hinsicht: Einerseits hinsichtlich der Kenngrößen

$$\begin{aligned} \text{Bias:} & \quad \frac{1}{l} \sum_{i=1}^l \hat{\rho}_i - \rho, \\ \text{Varianz:} & \quad \frac{1}{l} \sum_{i=1}^l (\hat{\rho}_i - \bar{\rho})^2 \\ \text{und MSE:} & \quad \frac{1}{l} \sum_{i=1}^l (\hat{\rho}_i - \rho)^2. \end{aligned}$$

Sie charakterisieren die Qualität der Schätzer (die Kenngrößen für σ^2 sind entsprechend definiert). Bei einer quadratischen Verlustfunktion wäre etwa der Schätzer mit dem kleineren mittleren quadratischen Fehler vorzuziehen. Andererseits interessieren die Schätzungen für ρ und σ^2 auch deshalb, weil aus ihnen ein Schätzer für die Kovarianzmatrix des KQ-Schätzers $\hat{\beta}$ gewonnen werden kann, indem man in der Formel

$$\text{Cov}(\hat{\beta}) = \sigma^2 (X^T X)^{-1} X^T (I - \rho W)^{-1} (I - \rho W^T)^{-1} X (X^T X)^{-1} \quad (23)$$

ρ und σ^2 durch die geschätzten Werte ersetzt. Damit können asymptotisch gültige Tests und Konfidenzbereiche für den Parametervektor β oder einzelne seiner Komponenten durchgeführt werden. Soll etwa getestet werden, ob β_j ,

die j -te Komponente von β , signifikant von einem unterstellten Wert β_{j0} verschieden ist, so ist die Nullhypothese $\beta_j = \beta_{j0}$ zu verwerfen, falls der Betrag der Differenz $\hat{\beta}_j - \beta_{j0}$, dividiert durch die Wurzel aus dem j -ten Hauptdiagonaleintrag von (23), größer ist als das $(1 - \frac{\alpha}{2})$ -Quantil der t -Verteilung mit $n - k$ Freiheitsgraden:

$$\text{Lehne die Nullhypothese ab, falls } \left| \frac{\hat{\beta}_j - \beta_{j0}}{\sqrt{\widehat{\text{Cov}}(\hat{\beta})_{jj}}} \right| \geq t_{n-k, 1-\frac{\alpha}{2}}.$$

Führt man diese Prozedur für jede der 10000 Wiederholungen aus, so ist der Anteil der 10000 Wiederholungen, bei denen die Nullhypothese verworfen wird, eine gute Näherung für das tatsächliche Niveau der Testprozedur. Dabei müssen keine verschiedenen Werte des Parametervektors β verwendet werden, weil der Schätzfehler nicht vom wahren Wert abhängt:

$$\hat{\beta} - \beta = (X^T X)^{-1} X^T (X\beta + u) - \beta = (X^T X)^{-1} X^T u.$$

Die Simulation benötigt keinen festen Wert für β und liefert dennoch Aussagen, die gleichzeitig für alle Werte von β gültig sind. In Bezug auf den Parameterraum für β gestaltet sich die Simulation somit komfortabler als in Bezug auf die Verwandtschaftsmatrix W und die Modellmatrix X - für die beiden letztgenannten müssen jeweils typische Vertreter ausgewählt werden.

Die Tabellen 1 bis 4 zeigen die Ergebnisse der Simulationen für die verschiedenen Werte von ρ . Die obere der jeweils zwei Zeilen betrifft die von Kelejian und Prucha (1999) vorgeschlagenen Schätzer; die untere Zeile enthält die Simulationsergebnisse für die in Kapitel 3 beschriebene Modifikation. Der Aufbau der vier Tabellen ist gleich: In den ersten drei Spalten stehen Bias, Varianz und mittlerer quadratischer Fehler der Schätzer für ρ , in den nächsten drei Spalten die entsprechenden Kenngrößen der Schätzer für σ^2 .

Tabelle 1: Kennzahlen der Schätzer für $n = 20$
 $\sigma^2 = 1$, $\alpha = 0,05$ und jeweils 10000 Wiederholungen;
 obere Zeilen: Werte für $\hat{\rho}_{NLS}$ und $\hat{\sigma}_{NLS}^2$;
 untere Zeilen: Werte für $\hat{\rho}_{NLS}^*$ und $\hat{\sigma}_{NLS}^{2*}$

$n = 20$ ρ	Schätzer für ρ			Schätzer für σ^2			$P\left(T_i > t_{n-k, 1-\frac{\alpha}{2}}\right)$		
	Bias	Var	MSE	Bias	Var	MSE	T_1	T_2	T_3
-0.8	-0.503	0.610	0.863	-0.267	0.086	0.157	0.049	0.067	0.048
	-0.113	0.520	0.533	-0.086	0.137	0.144	0.039	0.047	0.056
-0.6	-0.565	0.661	0.980	-0.267	0.084	0.155	0.047	0.079	0.044
	-0.133	0.551	0.569	-0.082	0.133	0.140	0.037	0.052	0.056
-0.4	-0.627	0.693	1.087	-0.281	0.081	0.160	0.053	0.093	0.044
	-0.158	0.588	0.613	-0.100	0.127	0.137	0.037	0.062	0.055
-0.2	-0.647	0.659	1.077	-0.273	0.080	0.154	0.065	0.119	0.042
	-0.157	0.547	0.571	-0.094	0.123	0.131	0.043	0.071	0.055
0	-0.683	0.640	1.107	-0.273	0.079	0.154	0.084	0.144	0.042
	-0.181	0.555	0.588	-0.100	0.119	0.129	0.053	0.084	0.054
0.2	-0.699	0.664	1.153	-0.259	0.082	0.149	0.107	0.179	0.042
	-0.174	0.556	0.586	-0.092	0.119	0.127	0.065	0.097	0.054
0.4	-0.680	0.594	1.057	-0.241	0.084	0.142	0.158	0.223	0.041
	-0.178	0.519	0.551	-0.082	0.120	0.127	0.092	0.117	0.053
0.6	-0.640	0.504	0.914	-0.214	0.089	0.135	0.246	0.264	0.039
	-0.144	0.474	0.495	-0.066	0.122	0.127	0.132	0.136	0.049
0.8	-0.590	0.425	0.773	-0.144	0.112	0.133	0.391	0.319	0.037
	-0.113	0.434	0.447	-0.005	0.146	0.146	0.215	0.147	0.048

Tabelle 2: Kennzahlen der Schätzer für $n = 60$
 $\sigma^2 = 1$, $\alpha = 0,05$ und jeweils 10000 Wiederholungen;
 obere Zeilen: Werte für $\hat{\rho}_{NLS}$ und $\hat{\sigma}_{NLS}^2$;
 untere Zeilen: Werte für $\hat{\rho}_{NLS}^*$ und $\hat{\sigma}_{NLS}^{2*}$

$n = 60$ ρ	Schätzer für ρ			Schätzer für σ^2			$P(T_i > t_{n-k, 1-\frac{\alpha}{2}})$		
	Bias	Var	MSE	Bias	Var	MSE	T_1	T_2	T_3
-0.8	-0.159	0.098	0.123	-0.104	0.035	0.046	0.048	0.066	0.048
	-0.048	0.096	0.099	-0.032	0.041	0.042	0.046	0.053	0.050
-0.6	-0.165	0.099	0.126	-0.101	0.034	0.044	0.051	0.078	0.051
	-0.046	0.095	0.097	-0.033	0.038	0.040	0.046	0.062	0.053
-0.4	-0.173	0.093	0.123	-0.091	0.033	0.042	0.053	0.084	0.047
	-0.052	0.088	0.091	-0.026	0.038	0.038	0.046	0.062	0.050
-0.2	-0.173	0.090	0.120	-0.090	0.032	0.040	0.058	0.096	0.048
	-0.054	0.084	0.087	-0.029	0.036	0.037	0.048	0.069	0.050
0	-0.171	0.083	0.112	-0.082	0.032	0.038	0.074	0.106	0.045
	-0.056	0.075	0.079	-0.027	0.035	0.036	0.058	0.075	0.047
0.2	-0.159	0.070	0.095	-0.074	0.032	0.037	0.081	0.105	0.048
	-0.052	0.062	0.065	-0.025	0.035	0.036	0.059	0.075	0.050
0.4	-0.144	0.055	0.076	-0.066	0.032	0.036	0.100	0.123	0.048
	-0.049	0.049	0.051	-0.023	0.035	0.035	0.071	0.081	0.048
0.6	-0.121	0.037	0.052	-0.047	0.033	0.036	0.123	0.126	0.050
	-0.039	0.035	0.037	-0.011	0.037	0.037	0.077	0.086	0.051
0.8	-0.083	0.026	0.033	-0.012	0.039	0.040	0.156	0.156	0.047
	-0.013	0.029	0.030	0.015	0.042	0.042	0.100	0.092	0.048

Tabelle 3: Kennzahlen der Schätzer für $n = 100$
 $\sigma^2 = 1$, $\alpha = 0,05$ und jeweils 10000 Wiederholungen;
 obere Zeilen: Werte für $\hat{\rho}_{NLS}$ und $\hat{\sigma}_{NLS}^2$;
 untere Zeilen: Werte für $\hat{\rho}_{NLS}^*$ und $\hat{\sigma}_{NLS}^{2*}$

$n = 100$ ρ	Schätzer für ρ			Schätzer für σ^2			$P\left(T_i > t_{n-k, 1-\frac{\alpha}{2}}\right)$		
	Bias	Var	MSE	Bias	Var	MSE	T_1	T_2	T_3
-0.8	-0.098	0.055	0.064	-0.064	0.022	0.026	0.050	0.069	0.048
	-0.029	0.054	0.055	-0.019	0.024	0.024	0.049	0.056	0.051
-0.6	-0.096	0.054	0.063	-0.061	0.021	0.025	0.052	0.072	0.050
	-0.025	0.053	0.054	-0.018	0.023	0.023	0.049	0.057	0.052
-0.4	-0.099	0.053	0.062	-0.058	0.020	0.024	0.052	0.071	0.049
	-0.029	0.051	0.052	-0.019	0.022	0.022	0.046	0.054	0.050
-0.2	-0.102	0.048	0.058	-0.053	0.019	0.022	0.059	0.075	0.049
	-0.033	0.046	0.047	-0.018	0.021	0.021	0.052	0.059	0.050
0	-0.095	0.040	0.049	-0.049	0.019	0.021	0.065	0.076	0.050
	-0.030	0.037	0.038	-0.017	0.020	0.020	0.054	0.061	0.051
0.2	-0.088	0.033	0.041	-0.045	0.019	0.021	0.072	0.086	0.048
	-0.029	0.031	0.032	-0.017	0.020	0.020	0.057	0.068	0.049
0.4	-0.079	0.025	0.031	-0.034	0.020	0.021	0.080	0.089	0.051
	-0.028	0.023	0.023	-0.009	0.021	0.021	0.062	0.068	0.052
0.6	-0.064	0.016	0.020	-0.029	0.020	0.021	0.094	0.100	0.050
	-0.022	0.015	0.016	-0.009	0.021	0.021	0.069	0.073	0.051
0.8	-0.038	0.010	0.012	-0.008	0.023	0.023	0.110	0.110	0.050
	-0.003	0.014	0.014	0.008	0.025	0.025	0.079	0.076	0.050

Tabelle 4: Kennzahlen der Schätzer für $n = 400$
 $\sigma^2 = 1$, $\alpha = 0,05$ und jeweils 10000 Wiederholungen
 obere Zeilen: Werte für $\hat{\rho}_{NLS}$ und $\hat{\sigma}_{NLS}^2$
 untere Zeilen: Werte für $\hat{\rho}_{NLS}^*$ und $\hat{\sigma}_{NLS}^{2*}$

$n = 400$ ρ	Schätzer für ρ			Schätzer für σ^2			$P\left(T_i > t_{n-k, 1-\frac{\alpha}{2}}\right)$		
	Bias	Var	MSE	Bias	Var	MSE	T_1	T_2	T_3
-0.8	-0.023	0.012	0.013	-0.017	0.006	0.006	0.049	0.054	0.048
	-0.005	0.012	0.012	-0.005	0.006	0.006	0.048	0.050	0.048
-0.6	-0.024	0.012	0.013	-0.015	0.005	0.005	0.051	0.055	0.053
	-0.006	0.012	0.012	-0.004	0.005	0.005	0.050	0.051	0.054
-0.4	-0.025	0.011	0.012	-0.015	0.005	0.005	0.051	0.054	0.052
	-0.007	0.011	0.011	-0.005	0.005	0.005	0.049	0.050	0.052
-0.2	-0.024	0.010	0.010	-0.014	0.005	0.005	0.051	0.055	0.051
	-0.007	0.010	0.010	-0.005	0.005	0.005	0.049	0.052	0.051
0	0.025	0.008	0.009	-0.012	0.005	0.005	0.057	0.055	0.048
	-0.010	0.008	0.008	-0.005	0.005	0.005	0.054	0.050	0.048
0.2	-0.021	0.006	0.007	-0.010	0.005	0.005	0.056	0.063	0.054
	-0.008	0.006	0.006	-0.003	0.005	0.005	0.052	0.058	0.054
0.4	-0.018	0.005	0.005	-0.009	0.005	0.005	0.053	0.064	0.048
	-0.006	0.005	0.005	-0.003	0.005	0.005	0.049	0.059	0.048
0.6	-0.013	0.003	0.003	-0.007	0.005	0.005	0.062	0.060	0.053
	-0.005	0.003	0.003	-0.002	0.005	0.005	0.056	0.055	0.053
0.8	-0.002	0.005	0.005	-0.004	0.005	0.005	0.067	0.065	0.051
	0.004	0.005	0.005	-0.000	0.005	0.005	0.061	0.058	0.052

In den letzten drei Spalten schließlich stehen die simulierten tatsächlichen Niveaus der oben beschriebenen t -Tests für die Regressionskoeffizienten β_1 , β_2 und β_3 (die ausgewählte Modellmatrix X hat drei Spalten). Das dabei vorgegebene Niveau ist $\alpha = 0,05$. Die jeweiligen Teststatistiken werden also mit dem 0,975-Quantil der t_{n-k} -Verteilung verglichen. Die erste Spalte von Tabelle 1 zeigt, dass beide Schätzer den Parameter ρ bei einem geringen Stichprobenumfang von $n = 20$ systematisch unterschätzen. Für den modifizierten Schätzer ist die Verzerrung allerdings wesentlich weniger gravierend: Sie beträgt nur etwa 20 – 25% der Verzerrung von $\hat{\rho}_{NLS}$ - der größte Teil der Verzerrung kann durch die Modifikation vermieden werden. Aus der zweiten Spalte ist ersichtlich, dass auch die Varianz deutlich zurückgeht, wenn auch nicht in einem so starken Ausmaße. Der mittlere quadratische Fehler verbessert sich durch die Modifikation somit auf etwa 50 – 60% des ursprünglichen Wertes.

Bei den Schätzern für σ^2 reduziert die Modifikation die Verzerrung ebenfalls sehr stark auf etwa ein Drittel des ursprünglichen Wertes. Im Gegensatz zur Situation bei der Schätzung von ρ steigt hier die Varianz an. Eine mögliche Erklärung für die gestiegene Varianz liefern die Matrizen G in (6) bzw. H in (15): Die jeweiligen dritten Spalten dieser Matrizen treffen bei der Multiplikation mit dem Vektor $(\rho, \rho^2, \sigma^2)^T$ auf den Varianzparameter σ^2 . Das letzte Element dieser Spalte ist bei der von Kelejian und Prucha (1999) verwendeten Matrix G gleich 0; in H lautet der entsprechende Eintrag $\frac{1}{n}\text{tr}(WM)$ und ist somit insbesondere von Null verschieden. Dadurch taucht σ^2 bei Verwendung von G in der zu minimierenden Funktion zweimal auf, bei Verwendung von H dreimal. Dies könnte die Ursache für den Varianzanstieg sein. Die größere Varianz wird jedoch durch die Reduktion der Verzerrung überkompensiert, so dass der mittlere quadratische Fehler dennoch geringer ausfällt - auch wenn die Verbesserung weniger deutlich ist als bei der Schätzung von ρ .

Die letzten drei Spalten zeigen das (simulierte) tatsächliche Niveau der

Signifikanztests für die Regressionskoeffizienten β_1 , β_2 bzw. β_3 bei einem vorgegebenen Signifikanzniveau von $\alpha = 0,05$. Es fallen deutliche Unterschiede zwischen den einzelnen Regressionskoeffizienten ins Auge: Während die Werte bei β_3 für alle ρ dem gewünschten Wert von 0,05 recht nahe kommen, gibt es bei den Tests für β_1 und β_2 deutliche Abweichungen vom vorgegebenen Niveau - es wird mitunter weit überschritten. Dies ist vor allem für stark positive Werte von ρ der Fall; für negative ρ gibt es nur bei β_2 leichte Überschreitungen, die mit wachsendem ρ zunehmen.

Positiv ist zu bemerken, dass die Niveauüberschreitungen durch die Modifikation des Schätzers entscheidend verringert werden: Der Anteil der Verwerfungen der Nullhypothese, welcher 5% übersteigt, kann um die Hälfte bis zwei Drittel vermindert werden. Die Wahrscheinlichkeit eines Fehlers erster Art liegt dadurch wesentlich näher am bewusst in Kauf genommenen Wert von 5% als vorher. Da das neue Verfahren den mittleren quadratischen Fehler bei der Schätzung der Parameter verringert, stellt die Verbesserung bei den Parametertests keine große Überraschung dar: Eine bessere Schätzung der Parameter führt zu einer besseren Schätzung der Kovarianzmatrix von $\hat{\beta}$, so dass das tatsächliche Niveau dichter am vorgegebenen Wert liegt. Das Ausmaß der Verbesserung ist jedoch schon etwas überraschend. Hier könnte eine Rolle spielen, dass die Verringerung des mittleren quadratischen Fehlers in erster Linie auf die wesentlich weniger verzerrte Schätzung der Parameter und erst in zweiter Linie auf eine Reduktion der Varianz zurückzuführen ist, weil dadurch auch der Schätzer für die Kovarianzmatrix von $\hat{\beta}$ eine geringere Verzerrung aufweisen dürfte. Es erscheint zumindest plausibel, dass dadurch das tatsächliche Niveau eines Parametertests näher am vorgegebenen Wert liegt als bei Verwendung einer Kovarianzmatrix, die aus stark verzerrten Schätzern für ρ und σ^2 gebildet wurde.

Die Tabellen 2 bis 4 zeigen die Ergebnisse der Simulation für die Stichprobenumfänge 60, 100 sowie 400. Erwartungsgemäß gehen Verzerrung und Varianz der Schätzer mit steigendem Stichprobenumfang zurück. Im Ver-

gleich der beiden Schätzverfahren untereinander bleiben die oben für $n = 20$ beschriebenen Erkenntnisse auch für die größeren Stichprobenumfänge erhalten: Durch die Modifikation der Schätzer geht die Verzerrung der Schätzer sehr stark zurück, die Varianz des Schätzers für ρ nimmt ebenfalls ab, wohingegen die Varianz des Schätzers für σ^2 leicht ansteigt. Insgesamt kann durch die Modifikation der mittlere quadratische Fehler reduziert werden. Der positive Einfluss auf das tatsächliche Niveau der Tests auf die Regressionsparameter bleibt ebenfalls erhalten: Die Niveauüberschreitungen gehen mit wachsendem Stichprobenumfang generell zurück; nichtsdestotrotz führt die Verwendung der modifizierten Schätzer selbst für $n = 400$ noch zu einer spürbaren Verbesserung der simulierten Niveaus. Eine Verwendung der modifizierten Schätzer lohnt sich demzufolge also nicht nur für moderate, sondern auch für größere Stichprobenumfänge.

Die Simulation dient in erster Linie einem Vergleich der beiden Schätzverfahren untereinander. Wie bereits oben kurz angedeutet, offenbaren die Tabellen 1 bis 4 jedoch noch weitere interessante Zusammenhänge. Einerseits unterscheiden sich die Ergebnisse der Signifikanztests für β_1 , β_2 und β_3 deutlich voneinander. Andererseits hängen die Resultate auch stark davon ab, wie groß ρ ist.

Die in den drei jeweils letzten Spalten der Tabellen 1 bis 4 aufgeführten simulierten Wahrscheinlichkeiten für einen Fehler erster Art liegen für β_3 in etwa auf dem vorgegebenen Niveau von 5%; für die anderen beiden Regressionskoeffizienten gibt es teilweise starke Überschreitungen dieser Grenze. Die Überschreitungen sind bei β_2 gravierender als bei β_1 . Dies dürfte mit der hier gewählten Struktur von X und W zusammenhängen: Die zweite Spalte von X besteht in der ersten Hälfte aus Einsen und in der zweiten Hälfte aus Nullen. In der dritten Spalte hingegen wechseln sich die Einträge Eins und Null ab. Jede einzelne Komponente des Störvektors wird per Konstruktion von den drei vorangehenden und den drei nachfolgenden Komponenten direkt beeinflusst. Dadurch gibt es unter den zur gleichen Stufe eines Faktors

gehörenden Komponenten in der zweiten Spalte stärkere Abhängigkeiten als in der dritten.

Sowohl die Qualität der Schätzung von ρ als auch die simulierten Ablehnungswahrscheinlichkeiten hängen zusätzlich davon ab, welchen Wert das wahre ρ annimmt. Die durchgezogenen Linien in Abbildung 2 zeigen den mittleren quadratischen Fehler der Schätzer für ρ als Funktion von ρ exemplarisch bei einem Stichprobenumfang von $n = 60$. Zunächst einmal ist deutlich zu erkennen, dass der mittlere quadratische Fehler des neuen Schätzers (schwarze Linie) wesentlich niedriger ist als der des alten Schätzers (graue Linie). Weiterhin fällt auf, dass die mittleren quadratischen Fehler für positive Werte von ρ weitaus geringer sind als bei negativen Werten von ρ . Für $\rho = 0,8$ sind die mittleren quadratischen Fehler bei beiden Schätzern etwa um den Faktor drei kleiner als für $\rho = -0,8$. Interessanterweise führt die bessere Schätzung des Parameters jedoch nicht dazu, dass das tatsächliche Niveau der Parametertests für die β_i dichter am vorgegebenen Wert von 0,05 liegt. Wie die gestrichelten Linien in Abbildung 2 verdeutlichen, ist sogar das Gegenteil der Fall. Hier sind die simulierten tatsächlichen Niveaus für die Tests auf β_2 abgetragen. Es ist deutlich zu erkennen, dass die Niveauüberschreitungen bei Verwendung des alten Schätzers (graue gestrichelte Linie) durch die Verwendung des neuen Schätzers (schwarz gestrichelt) entscheidend reduziert werden. Die Überschreitungen des vorgegebenen Signifikanzniveaus sind darüber hinaus für positive Werte von ρ wesentlich gravierender als für negative. Obwohl positive ρ besser geschätzt werden können als negative, wird das vorgegebene Signifikanzniveau dennoch stärker überschritten.

Diese auf den ersten Blick widersprüchliche Beobachtung lässt sich unter Verwendung von Matrixdifferentiation analytisch wie folgt erklären. Die Einträge der Störgrößenkovarianzmatrix ($\sigma^2 = 1$)

$$\text{Cov}(u) = (I - \rho W)^{-1}(I - \rho W^T)^{-1}$$

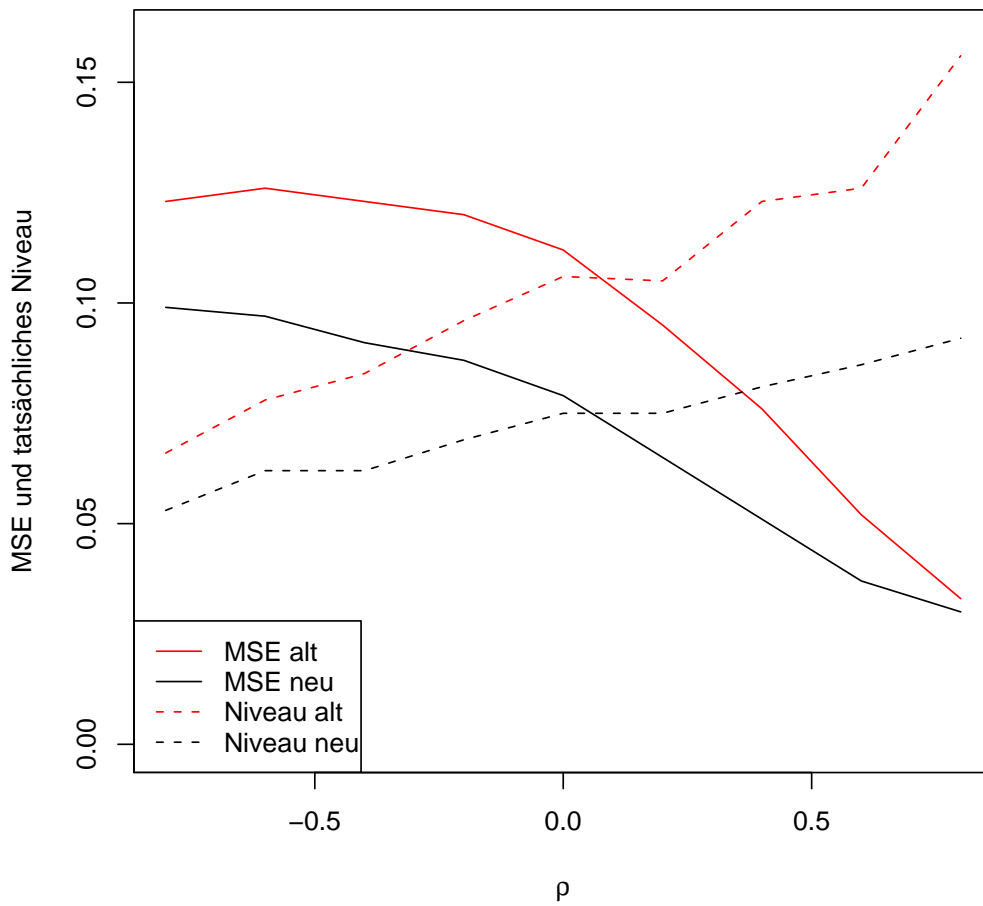


Abbildung 2: Einfluss des räumlichen Autokorrelationsparameters bei einem Stichprobenumfang von $n = 60$ und einem vorgegebenen Niveau von $\alpha = 0,05$ für den alten GMM-Schätzer (graue Linien) sowie die Modifikation (schwarze Linien)

reagieren für positive ρ auf kleine Änderungen von ρ deutlich stärker als für negative ρ . Nach Theorem 3 auf S. 151 in Magnus und Neudecker (1988) gilt nämlich

$$\begin{aligned}\frac{\partial(I - \rho W)^{-1}}{\partial \rho} &= -(I - \rho W)^{-1} \frac{\partial(I - \rho W)}{\partial W} (I - \rho W)^{-1} \\ &= (I - \rho W)^{-1} W (I - \rho W)^{-1}\end{aligned}$$

sowie

$$\frac{\partial(I - \rho W^T)^{-1}}{\partial \rho} = (I - \rho W^T)^{-1} W^T (I - \rho W^T)^{-1},$$

so dass die Ableitung von $\text{Cov}(u)$ nach ρ durch

$$\begin{aligned}\frac{\partial(I - \rho W)^{-1} (I - \rho W^T)^{-1}}{\partial \rho} &= (I - \rho W)^{-1} (I - \rho W^T)^{-1} W^T (I - \rho W^T)^{-1} \\ &\quad + (I - \rho W^T)^{-1} (I - \rho W)^{-1} W (I - \rho W)^{-1}\end{aligned}$$

gegeben ist. Die Ableitung von

$$\text{Cov}(\hat{\beta}) = (X^T X)^{-1} X^T (I - \rho W)^{-1} (I - \rho W^T)^{-1} X (X^T X)^{-1} \quad (24)$$

nach ρ lautet somit

$$\begin{aligned}\frac{\partial \text{Cov}(\hat{\beta})}{\partial \rho} &= (X^T X)^{-1} X^T \left[(I - \rho W)^{-1} (I - \rho W^T)^{-1} W^T (I - \rho W^T)^{-1} \right. \\ &\quad \left. + (I - \rho W^T)^{-1} (I - \rho W)^{-1} W (I - \rho W)^{-1} \right] X (X^T X)^{-1}.\end{aligned}$$

Die Einträge dieser Matrix unterscheiden sich stark danach, ob ρ positiv oder

negativ ist. So lautet die Matrix etwa für $\rho = -0,8$

$$\left. \frac{\partial \text{Cov}(\hat{\beta})}{\partial \rho} \right|_{\rho=-0,8} = \begin{pmatrix} -0,018 & -0,009 & 0,056 \\ -0,009 & 0,018 & 0 \\ 0,056 & 0 & -0,013 \end{pmatrix},$$

während sich für $\rho = 0,8$

$$\left. \frac{\partial \text{Cov}(\hat{\beta})}{\partial \rho} \right|_{\rho=0,8} = \begin{pmatrix} 6,67 & -5,02 & 0,01 \\ -5,02 & 10,04 & 0 \\ 0,01 & 0 & -0,02 \end{pmatrix}$$

ergibt. Die Varianz von $\hat{\beta}_2$ reagiert an der Stelle $\rho = 0,8$ um ein Vielfaches stärker auf kleine Veränderungen von ρ als an der Stelle $\rho = -0,8$. Diese Veränderungen unterscheiden sich nicht nur absolut, sondern auch relativ. Die in (24) angegebene Kovarianzmatrix von $\hat{\beta}$ ist für $\rho = -0,8$ gleich

$$\text{Cov}(\hat{\beta}) \Big|_{\rho=-0,8} = \begin{pmatrix} 0,043 & -0,013 & -0,062 \\ -0,013 & 0,027 & 0 \\ -0,062 & 0 & 0,113 \end{pmatrix}$$

und für $\rho = 0,8$ gleich

$$\text{Cov}(\hat{\beta}) \Big|_{\rho=0,8} = \begin{pmatrix} 0,719 & -0,583 & -0,021 \\ -0,583 & 1,166 & 0 \\ -0,021 & 0 & 0,042 \end{pmatrix}.$$

Damit bewirken minimale Änderungen von ρ an der Stelle $\rho = -0,8$ eine relative Änderung der Varianz von $\hat{\beta}_2$ in Höhe von $\frac{0,018}{0,027} = \frac{2}{3}$; an der Stelle $\rho = 0,8$ beträgt dieser Wert 8,61. Das erklärt Abbildung 2: Obwohl positive Werte von ρ besser geschätzt werden können als negative, kann die Varianz von $\hat{\beta}_2$ schlechter geschätzt werden, so dass sich die Teststatistik des Signifikanztests für β_2 häufiger im Ablehnungsbereich realisiert.

Diese relativen Änderungen der Varianzen der Regressionskoeffizienten bei minimalen Änderungen von ρ erklären auch nochmals die unterschiedlichen Ablehnungswahrscheinlichkeiten der Tests auf β_1 und β_3 : Für β_1 betragen sie $-0,42$ an der Stelle $\rho = -0,8$ und $9,28$ an der Stelle $\rho = 0,8$; dies führt zu deutlichen Niveauüberschreitungen bei stark positiven Werten von ρ , während für negative Werte von ρ das vorgegebene Niveau in etwa eingehalten wird. Für β_3 wird das vorgegebene Niveau für alle ρ in etwa eingehalten. Passend dazu betragen die relativen Änderungen $-0,12$ an der Stelle $\rho = -0,8$ und $-0,48$ an der Stelle $\rho = 0,8$.

7 Anwendung auf Arbeitslosendaten

Seit vielen Jahren gilt die Arbeitslosigkeit als eines der größten Probleme in Deutschland. Dabei sind nicht alle Bevölkerungsgruppen in gleicher Weise betroffen. Stattdessen unterscheiden sich die Arbeitslosenquoten in Teilgruppen der Bevölkerung stark voneinander (vgl. Kapitel 9.1 in Franz (2003)). Solche Unterschiede können zwischen Regionen (höhere Arbeitslosenquoten im Osten Deutschlands), Bildungsgraden (höhere Arbeitslosenquoten für gering Qualifizierte), Altersklassen (höhere Arbeitslosenquoten für ältere oder jüngere Personen) oder auch zwischen Männern und Frauen beobachtet werden. Saisonale Effekte (höhere Arbeitslosenquoten in den Wintermonaten) sowie die allgemeine Konjunktur beeinflussen das Ausmaß der Arbeitslosigkeit zusätzlich. Zur Bekämpfung der Arbeitslosigkeit hat die Politik verschiedene Maßnahmen entwickelt. Diese Maßnahmen sollen möglichst effizient eingesetzt werden, indem sie etwa auf einzelne Bevölkerungsgruppen zugeschnitten sind. Der gezielte Einsatz solcher Maßnahmen setzt zunächst eine genaue Kenntnis über die Struktur der Arbeitslosigkeit voraus. Dieses Kapitel möchte hierzu einen Beitrag leisten, indem der in Kapitel 3 vorgestellte Schätzer auf Daten zur Arbeitslosigkeit in Deutschland angewendet wird.

7.1 Die Daten

Der Datensatz stammt aus dem sozio-ökonomischen Panel. Er enthält insgesamt 1.449.355 Beobachtungspunkte für die alten Bundesländer sowie 433.373 Beobachtungspunkte für die neuen Bundesländer. Jeder Datenpunkt gibt für eine Person den Beschäftigungsstatus in einem bestimmten Monat an. Dabei wird zwischen den drei möglichen Ausprägungen „beschäftigt“, „arbeitslos“ sowie „nicht partizipierend“ unterschieden. Als „beschäftigt“ gilt, wer entweder voll- oder teilzeit beschäftigt ist. Auch Teilnehmer an beruflichen Fortbildungsmaßnahmen fallen in diese Kategorie. Die Kategorie „arbeitslos“ umfasst die registrierten Arbeitslosen, die keiner Erwerbsarbeit nachge-

hen. Als „nicht partizipierend“ gilt jeder, der nicht in eine der beiden anderen Kategorien eingeordnet werden kann. Es handelt sich beispielsweise um Schüler, Studenten, Wehr- oder Zivildienstleistende, Personen, die sich um Kinder oder pflegebedürftige Angehörige kümmern, oder auch Frührentner. Zusätzlich zum Beschäftigungsstatus und der Angabe von Monat und Jahr enthält jeder Datenpunkt vier weitere Informationen über die jeweilige Person: Das Geschlecht, das Alter, den höchsten erreichten Bildungsabschluss und das Gewicht des Datenpunktes im sozio-ökonomischen Panel.

Das Alter ist in drei Klassen angegeben, 16-24, 25-49 sowie 50-64 Jahre. Der Bildungsabschluss wird ebenfalls in drei Kategorien gemessen. Als „hoch“ gilt dabei ein Hochschul- oder Fachhochschulabschluss, als „mittel“ das Abitur oder die mittlere Reife mit anschließender Berufsausbildung. Personen ohne solche Abschlüsse fallen in die Kategorie „niedrig“. Die unterschiedliche Gewichtung der Personen im sozio-ökonomischen Panel ergibt sich daraus, dass die Wahrscheinlichkeit, in diese Erhebung eingeschlossen zu werden, nicht für alle Personen gleich ist. Das wiederum liegt daran, dass das sozio-ökonomische Panel verlässliche Auswertungen auch für Teilgruppen der Bevölkerung ermöglichen soll; um in diesen Teilgruppen eine hinreichende Stichprobengröße zu erzielen, werden Angehörige solcher Teilgruppe überproportional in die Erhebung aufgenommen. Das ist beispielsweise der Fall für ausländische Mitbürger. Der Datensatz enthält für jeden Beobachtungspunkt zwei solcher Gewichte - eines für Querschnittsuntersuchungen und eines für Paneluntersuchungen.

Tabelle 5 zeigt die (ungewichtete) Verteilung der Merkmale in den Datensätzen getrennt nach Ost- und Westdeutschland. Während das Geschlecht sowie das Alter in beiden Teildatensätzen vergleichbare relative Häufigkeiten aufweisen, gibt es bei den Merkmalen Beschäftigungsstatus und höchster Bildungsabschluss deutliche Unterschiede: Der Anteil der Arbeitslosen ist in Ostdeutschland wesentlich größer als in Westdeutschland; gleichzeitig nimmt die Ausprägung „nicht partizipierend“ in Ostdeutschland einen geringeren

Tabelle 5: Überblick über den Datensatz
(relative Häufigkeiten und Stichprobengröße)

Merkmal	West	Ost
Frau	50,71	50,96
Mann	49,29	49,04
beschäftigt	59,50	63,56
arbeitslos	2,82	10,58
nicht partizipierend	37,69	25,86
Alter 16-24	17,63	17,28
Alter 25-49	57,11	55,27
Alter 50-64	25,26	27,46
hoher Bildungsabschluss	11,43	23,69
mittlerer Bildungsabschluss	64,89	62,87
niedriger Bildungsabschluss	23,68	13,44
Datenpunkte	1.449.355	433.373

Anteil ein. Dies dürfte an der in Ostdeutschland traditionell höheren Frauenerwerbsquote liegen.

Der Prozentsatz arbeitsloser Personen wirkt vergleichsweise niedrig; insbesondere fällt er niedriger aus als die monatlich von der Bundesagentur für Arbeit veröffentlichten Arbeitslosenquoten. Dieser scheinbare Unterschied ist auf mehrere Effekte zurückzuführen: Zunächst einmal gibt Tabelle 5 den Anteil der Arbeitslosen an allen in der Stichprobe enthaltenen Personen an; insbesondere tauchen bei der Anteilsberechnung im Nenner auch die nicht partizipierenden Personen auf. Diese spielen in den von der Bundesagentur veröffentlichten Arbeitslosenquoten keine Rolle, weil hier der Anteil der Arbeitslosen an den Erwerbspersonen angegeben wird. Als zweites sind die in Tabelle 5 angegebenen Werte nicht gewichtet, woraus eine zusätzliche Verzerrung resultiert. Und schließlich handelt es sich um zwei verschiedene Konzepte zur Ermittlung der Arbeitslosenquote. Während für die Statistik der Bundesagentur als arbeitslos gilt, wer sich bei der örtlichen Arbeitsagentur als arbeitsuchend gemeldet hat, handelt es sich bei dem sozio-ökonomischen Panel um eine Befragung der Personen. Dabei kann es durchaus vorkommen, dass offiziell als arbeitsuchend registrierte Personen sich selbst gar nicht als arbeitslos klassifizieren. Die Meldung bei der örtlichen Arbeitsagentur könnte vielmehr aus einem damit verbundenen Anspruch auf Transferzahlungen motiviert sein. Darüber hinaus werden manche Personengruppen in den beiden Berechnungskonzepten unterschiedlich behandelt, z.B. Teilzeitbeschäftigte, die weniger als 15 Stunden pro Woche arbeiten und gleichzeitig eine Vollzeitbeschäftigung suchen: Für die Statistik der Bundesagentur gelten diese Personen als arbeitslos, das sozio-ökonomische Panel zählt sie als „beschäftigt“.

Aus Tabelle 5 ist ersichtlich, dass alle Merkmalsausprägungen im Datensatz hinreichend oft vertreten sind. Das gilt jedoch nicht für jede Kombination zweier Merkmale. Tabelle 6 zeigt Kreuztabellen der Merkmale „Bildungsabschluss“ und „Alter“ getrennt für Westdeutschland (Teil (a) vor der Einheit, Teil (b) nach der Einheit) und Ostdeutschland (Teil (c)). Obwohl die Ge-

Tabelle 6: Verteilung der Stichprobe auf die demographischen Gruppen
 Absolutzahlen sowie Prozentzahlen pro Zeile,
 in der letzten Spalte Prozentzahlen pro Spalte

a) Bildungsabschluss und Alter in Westdeutschland vor der Einheit
 (Januar 1982 bis Juni 1990)

Alter	Bildungsabschluss						Σ	
	niedrig		mittel		hoch			
16-24	62375	(48%)	65520	(51%)	1118	(1%)	129013	(21%)
25-49	50656	(16%)	229872	(71%)	42719	(13%)	323247	(54%)
50-64	49714	(33%)	87782	(59%)	11724	(8%)	149220	(25%)
Σ	162745	(27%)	383174	(64%)	55561	(9%)	601480	(100%)

b) Bildungsabschluss und Alter in Westdeutschland nach der Einheit
 (Juli 1990 bis Dezember 2003)

Alter	Bildungsabschluss						Σ	
	niedrig		mittel		hoch			
16-24	63890	(50%)	61270	(48%)	1403	(1%)	126563	(15%)
25-49	65330	(13%)	356624	(71%)	82534	(16%)	504488	(60%)
50-64	51248	(24%)	139383	(64%)	26193	(12%)	216824	(26%)
Σ	180469	(21%)	557277	(66%)	110130	(13%)	847875	(100%)

c) Bildungsabschluss und Alter in Ostdeutschland
 (Januar 1992 bis Dezember 2003)

Alter	Bildungsabschluss						Σ	
	niedrig		mittel		hoch			
16-24	36480	(49%)	37033	(49%)	1358	(2%)	74871	(17%)
25-49	9559	(4%)	162832	(68%)	67127	(28%)	239518	(55%)
50-64	12215	(10%)	72603	(61%)	34166	(29%)	118984	(27%)
Σ	58254	(13%)	272468	(63%)	102651	(24%)	433373	(100%)

samtzahl der Beobachtungspunkte sehr groß ist, kommt eine Merkmalskombination vergleichsweise selten vor: Es gibt wenige junge Personen mit einem hohen Bildungsabschluss. Das erscheint plausibel, da nur wenige Hochschul­er ihren Abschluss vor dem 25. Lebensjahr erreichen. Neben den Absolutzahlen zeigt Tabelle 6 auch die Verteilung der Bildungsabschlüsse in den einzelnen Altersklassen. Hier fällt auf, dass der Anteil der Geringqualifizierten im Westen von 27% vor der Einheit auf 21% nach der Einheit gesunken ist. Gleichzeitig ist der Anteil der Hochqualifizierten von 9% auf 13% gestiegen. In Ostdeutschland ist das formale Bildungsniveau nochmals höher; der Anteil der Hochqualifizierten beträgt 24%, der der Geringqualifizierten nur 13%. In der mittleren Altersklasse ist der Anteil der Geringqualifizierten mit 4% besonders niedrig.

7.2 Das Modell

Die gruppenspezifische Arbeitslosenquote im Monat m des Jahres t erklärt sich in Anlehnung an Kluge et al. (2006) durch das Modell

$$\begin{aligned} \mu_{igmt} = & (\alpha + \gamma \cdot 1_{Frau}) + \sum_{i \neq 5} (\beta_i + \delta_i \cdot 1_{Frau}) + \sum_{m \neq 6} \lambda_m 1_m \\ & + \text{bip}_{t-1} \sum_{g=1}^{18} \theta_g \cdot 1_g + u_{igmt}. \end{aligned} \quad (25)$$

Der Index $i \in (1, \dots, 9)$ durchläuft die durch Alter und Bildungsabschluss definierten demographischen Gruppen; der Index $g \in (1, \dots, 18)$ steht für die 18 zusätzlich nach Geschlecht unterteilten demographischen Gruppen. Die verschiedenen Kalendermonate werden durch den Index $m \in (1, \dots, 12)$ gekennzeichnet. Die Indikatorvariablen 1_{Frau} , 1_m und 1_g nehmen den Wert 1 an, wenn die Arbeitslosenquote auf der linken Seite von (25) das entsprechende Merkmal aufweist, und sind ansonsten gleich 0. Die Summanden mit

$i = 5$ und $m = 6$ tauchen in der Modellgleichung nicht auf. Hierbei handelt es sich um die Referenzklassen, deren Effekte a priori auf Null gesetzt werden. Für die demographischen Gruppen ist dies die Klasse mit mittlerem Alter und mittlerem Bildungsabschluss; für die Kalendermonate ist es der Juni. Dementsprechend ist α die erwartete Arbeitslosenquote im Monat Juni für die männliche Gruppe mit mittlerem Alter und mittlerem Bildungsabschluss. Der Wert für die entsprechende weibliche Gruppe ist $\alpha + \gamma$. Die Parameter β_i stehen für die Differenz in den Arbeitslosenquoten der männlichen Gruppe i zur männlichen Referenzgruppe. Die erwartete Arbeitslosenquote für gut ausgebildete alte Männer ist somit beispielsweise $\alpha + \beta_9$. Die Unterschiede in den Differenzen der demographischen Gruppen zwischen Männern und Frauen werden von den Parametern δ_i aufgefangen; die erwartete Arbeitslosenquote für junge, schlecht ausgebildete Frauen ist beispielsweise $\alpha + \gamma + \beta_1 + \delta_1$. Dies gilt jeweils für den Referenzmonat Juni, die Monatseffekte schlagen sich in den Parametern λ_m nieder. Sie geben an, um wieviel die erwartete Arbeitslosenquote im Monat m höher ist als im Juni. Die Parameter θ_g schließlich fangen den Einfluss der Konjunktur auf die Arbeitslosigkeit auf, und zwar separat für alle 18 demographischen Gruppen. Die Wachstumsrate des Bruttoinlandsprodukts des Vorjahres, bip_{t-1} , steht für die aktuelle Konjunktur. Da sich konjunkturelle Einflüsse auf dem Arbeitsmarkt oftmals erst mit einiger Verzögerung niederschlagen, wird die Wachstumsrate des Vorjahres verwendet. Die erwartete Arbeitslosenquote in der demographischen Gruppe i bzw. g , Monat m und Jahr t setzt sich also für Männer zusammen aus der für die männliche Referenzgruppe erwarteten Quote, α , dem differentiellen Effekt der demographischen Gruppe, β_i , dem differentiellen Effekt des Kalendermonats im Vergleich zum Juni, λ_m , sowie dem Konjunkturreffekt für die demographische Gruppe g , θ_g , multipliziert mit dem Wirtschaftswachstum des Vorjahres, bip_{t-1} . Für weibliche Gruppen kommen noch die differentiellen Effekte γ sowie δ_i hinzu.

Wie immer in Regressionsmodellen gilt dieser Zusammenhang nicht per-

fekt. Stattdessen überlagert der Störterm u_{igmt} den linearen Zusammenhang. Die in Regressionsmodellen oft unterstellten Annahmen paarweise unkorrelierter und homoskedastischer Störterme erscheinen hier unrealistisch.

Einerseits werden die Störterme nicht unkorreliert sein: Wenn die Effekte über einen Zeitraum von mehreren Jahren geschätzt werden, dann dürfte das Arbeitslosigkeitsniveau im manchen Jahren systematisch über dem Durchschnitt und in anderen Jahren systematisch darunter liegen - und das gilt dann für alle demographischen Gruppen gleichermaßen. Auch die in manchen Monaten wirkenden Sondereinflüsse wie ein besonders milder oder harter Winter oder die Fußball-Weltmeisterschaft 2006 beeinflussen die Störgrößen in den entsprechenden Monaten tendenziell in die gleiche Richtung. Dadurch erscheint eine positive Korrelation von zeitlich dicht beieinander liegenden Störtermen plausibel.

Andererseits werden die Störterme unterschiedliche Varianzen aufweisen: Da die demographischen Gruppen unterschiedlich stark besetzt sind, berechnen sich die Arbeitslosenquoten μ_{igmt} aus unterschiedlich großen Gruppen von Personen, die zudem im sozio-ökonomischen Panel noch verschiedene Gewichte aufweisen. Zu großen Gruppen korrespondierende Störterme werden eine kleinere Varianz aufweisen als solche, die zu einer aus einer geringen Anzahl von Personen berechneten Arbeitslosenquote gehören.

Aus den soeben dargestellten Überlegungen heraus bietet es sich an, die Störgrößen wie in (1) als räumlich korreliert zu modellieren:

$$u = \rho W u + \varepsilon.$$

Die in der Verwandtschaftsmatrix W zusammengefassten direkten Einflüsse der gruppenspezifischen Arbeitslosenquoten haben folgende Struktur: Die Arbeitslosenquote einer demographischen Gruppe in einem bestimmten Monat ist direkt verwandt mit allen anderen Gruppen im gleichen Monat (das ergibt maximal 17 Nachbarn) sowie mit den Arbeitslosenquoten der gleichen

Tabelle 7: Anzahl direkter Nachbarn

	17	18	19
Westdeutschland (1983-1990)	0	358	1258
Westdeutschland (1990-2003)	122	1229	1508
Ostdeutschland (1992-2003)	37	1487	979

Gruppe in den beiden Monaten direkt vor bzw. direkt nach dem Beobachtungsmonat (das ergibt maximal zwei weitere Nachbarn). Insgesamt können in einer Zeile von W also maximal 19 Einträge von Null verschieden sein. Tabelle 7 zeigt, dass diese Höchstzahl nicht immer erreicht wird. Das liegt einerseits daran, dass in manchen Monaten nicht für alle demographischen Gruppen Arbeitslosenquoten vorliegen (hier ist vor allem an die Gruppe der jungen Hochqualifizierten zu denken) und andererseits daran, dass für den ersten bzw. letzten Monat des Beobachtungszeitraumes keine Daten im Monat zuvor bzw. danach vorliegen. Bei jeder einzelnen Beobachtung liegen aber zumindest für 17 der 19 möglichen Nachbarn Arbeitslosenquoten vor.

Die in diesem Modell üblicherweise getroffene Annahme $\text{Cov}(\varepsilon) = \sigma^2 I$ ist hier nicht realistisch: Da die verschiedenen Komponenten von u unterschiedliche Varianzen besitzen, muss dies auch für die verschiedenen Komponenten von ε gelten. Deshalb soll die Kovarianzmatrix von ε eine Diagonalmatrix sein, bei der sich die Einträge auf der Hauptdiagonalen voneinander unterscheiden dürfen,

$$\text{Cov}(\varepsilon) = \sigma^2 V \quad \text{mit} \quad V = \text{diag}(v_{11}, \dots, v_{nn}).$$

Die Matrix V wird standardisiert, so dass $\text{tr}(V) = n$ gilt. Dadurch wird die Schätzung für σ^2 eindeutig. Damit gilt

$$\text{Cov}(u) = \sigma^2(I - \rho W)^{-1}V(I - \rho W^T)^{-1} =: \Sigma$$

und die Schätzung des Modells erfolgt mit einer iterativen Prozedur. Im ersten Schritt wird ausgehend von Startwerten für ρ und σ^2 die Matrix Σ und damit der verallgemeinerte Kleinst-Quadrate-Schätzer

$$\tilde{\beta} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} y$$

berechnet. Daraus ergeben sich die Regressionsresiduen

$$\hat{u} = y - X(X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} y =: \tilde{M}y,$$

aus denen mit Hilfe des modifizierten GMM-Schätzers Schätzungen für ρ und σ^2 gewonnen werden.

Es bietet sich an, den GMM-Schätzer anstelle des Maximum-Likelihood-Schätzers zu verwenden, weil der Maximum-Likelihood-Schätzer bei den hier vorliegenden großen Stichprobenumfängen numerisch instabil ist. Trotz der großen Stichprobenumfänge lohnt es sich darüber hinaus, den modifizierten GMM-Schätzer zu verwenden: Einerseits erfordert er keinen zusätzlichen Rechenaufwand; andererseits zeigt die Simulation, dass er auch bei großen Stichprobenumfängen akurater Signifikanztests für die Regressionskoeffizienten ermöglicht.

Aus den auf diese Weise gewonnenen Schätzungen für ρ und σ^2 wird im zweiten Schritt die Matrix Σ und damit der verallgemeinerte Kleinst-Quadrate-Schätzer mit der neuen Matrix Σ erneut bestimmt. Aus den neuen Regressionsresiduen können wiederum neue Schätzungen für ρ und σ^2 gewonnen werden, mit denen eine weitere Iteration gestartet wird. Das Verfahren wird solange wiederholt, bis sich die Schätzungen für β , ρ und σ^2 nicht mehr wesentlich verändern. Aus den auf diese Weise erhaltenen Schätzungen für ρ

und σ^2 ergibt sich $\hat{\Sigma}$ als Schätzung für Σ . Daraus erhält man wiederum eine Schätzung für die Kovarianzmatrix von $\tilde{\beta}$ durch

$$\widehat{\text{Cov}}(\tilde{\beta}) = (X^T \hat{\Sigma}^{-1} X)^{-1},$$

mit deren Hilfe asymptotisch gültige Parametertests sowie Konfidenzbereiche für die Komponenten des Parametervektors β gewonnen werden können. Diese sind nur asymptotisch gültig, weil sie auf den Schätzern für ρ und σ^2 beruhen. Die verwendeten GMM-Schätzer sind nach Theorem 3.1 zwar konsistent, stimmen aber für einen endlichen Stichprobenumfang nicht exakt mit den wahren Werten überein.

7.3 Ergebnisse

Dieses Unterkapitel präsentiert die Ergebnisse der im vorangegangenen Unterkapitel vorgestellten Schätzprozedur. Um diese anwenden zu können, muss der Rohdatensatz aus dem sozio-ökonomischen Panel zunächst aufbereitet werden. Die gruppenspezifische Arbeitslosenquote in einem bestimmten Monat errechnet sich für jede demographische Gruppe als Anzahl der im entsprechenden Monat arbeitslosen Personen, dividiert durch die Summe der arbeitslosen sowie der beschäftigten Personen. Das Gewicht einer beobachteten Arbeitslosenquote ist die Summe der Querschnittsgewichte von denjenigen Personen, die im jeweiligen Monat zur betreffenden demographischen Gruppe gehören und entweder arbeitslos oder beschäftigt sind. Da der Panelcharakter der Rohdaten nicht direkt in die Auswertung einfließt, werden die Querschnittsgewichte und nicht etwa die Panelgewichte verwendet.

Das Modell wird auf drei Teildatensätzen geschätzt: Erstens für Ostdeutschland einschließlich Berlin (Januar 1992 - Dezember 2003), zweitens für Westdeutschland (einschließlich West-Berlin) vor der Einheit (Januar 1983 - Juni 1990) und drittens für Westdeutschland ohne Berlin nach der Einheit (Juli 1990 - Dezember 2003). Für den Zeitpunkt der Einheit wird nicht das

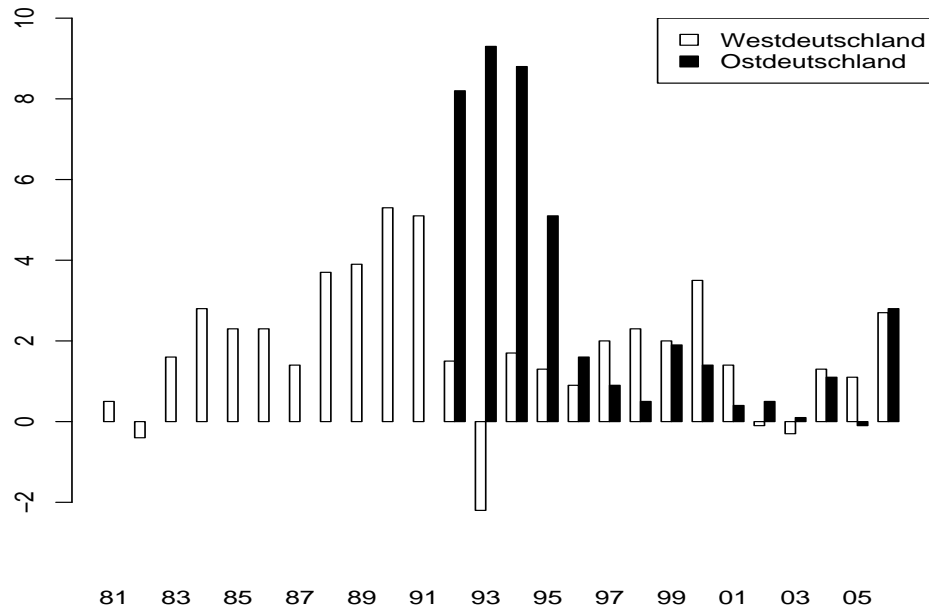


Abbildung 3: Reales jährliches Wachstum des Bruttoinlandsprodukts in Deutschland getrennt nach Westdeutschland (1981-2006) sowie Ostdeutschland (1992-2006)

Datum der politischen Einheit, sondern der Zeitpunkt der Wirtschafts- und Währungsunion zugrunde gelegt, die am 1.7.1990 in Kraft trat. Die Aufteilung des Gesamtdatensatzes ist erforderlich, weil die Annahme konstanter Effekte über den gesamten Datensatz fragwürdig erscheint. So existieren beispielsweise deutliche Unterschiede im Niveau der Arbeitslosigkeit zwischen Ost- und Westdeutschland. Auch innerhalb des westdeutschen Arbeitsmarktes dürften Unterschiede zwischen den Zeiträumen vor bzw. nach der deutschen Einheit bestehen.

Abbildung 3 stellt die jährlichen Wachstumsraten des Bruttoinlandspro-

duktes in Deutschland getrennt für Ost- und Westdeutschland dar. Auffällig sind die hohen Wachstumsraten in Ostdeutschland (schwarze Balken) in den ersten Jahren nach der Einheit. Die mit -2,2 sehr niedrige Rate für Westdeutschland (weiße Balken) im Jahr 1993 deutet darauf hin, dass in den ersten Jahren nach der Wiedervereinigung ein Teil der wirtschaftlichen Aktivitäten aus Westdeutschland nach Ostdeutschland verlagert wurde. Hohe Wachstumsraten von über drei Prozent finden sich in Westdeutschland im Zeitraum von 1988 bis 1991 sowie im Jahr 2000. In Ostdeutschland waren die Wachstumsraten in den ersten Jahren nach der Einheit sehr hoch; ab dem Jahr 1996 bleiben die ostdeutschen Wachstumsraten meist hinter den westdeutschen zurück. Da für das Jahr 1991 in Ostdeutschland aufgrund der fehlenden Ausgangsbasis im Jahr 1990 keine Wachstumsrate ermittelt werden kann, lege ich hier den Wert für Westdeutschland zugrunde. Die in Abbildung 3 dargestellten Wachstumsraten dienen zur Schätzung der Abhängigkeit der gruppenspezifischen Arbeitslosenquoten von der Konjunktur. Die Effekte θ_g ergeben sich aus der Regression der gruppenspezifischen Arbeitslosenquoten auf die Wachstumsraten des jeweiligen Vorjahres. Dazu wird die im letzten Unterkapitel verwendete iterative Prozedur verwendet.

Tabelle 8 gibt einen ersten Überblick über die Regression auf den drei Teildatensätzen. Neben der Anzahl der Beobachtungspunkte (Anzahl der Kombinationen aus Monat und demographischer Gruppe, für die eine Arbeitslosenquote ermittelt werden konnte) sind das korrigierte Bestimmtheitsmaß,

$$\bar{R}^2 = R^2 - \frac{k-1}{n-k}(1-R^2) \text{ mit } R^2 = \frac{(\hat{y} - \bar{y})^T \hat{\Sigma}^{-1} (\hat{y} - \bar{y})}{(y - \bar{y})^T \hat{\Sigma}^{-1} (y - \bar{y})},$$

sowie die Schätzungen für die Modellparameter ρ und σ^2 aufgeführt. Die Modellanpassung ist für Ostdeutschland mit einem korrigierten Bestimmtheitsmaß von über 90% besser als für die beiden westdeutschen Regressionen. Die Schätzungen für den räumlich autoregressiven Parameter ρ liegen dicht beieinander, sie betragen jeweils etwa 0,25. Bei der Standardabweichung der

Tabelle 8: Übersicht über die Regressionen

	n	\bar{R}^2	$\hat{\rho}$	$\hat{\sigma}$
Westdeutschland (1983-1990)	1616	0,753	0,279	0,031
Westdeutschland (1990-2003)	2859	0,785	0,244	0,042
Ostdeutschland (1992-2003)	2503	0,914	0,244	0,070

Innovationen hingegen gibt es deutliche Unterschiede: Während die Abweichungen für Westdeutschland moderat sind (etwa 3 Prozentpunkte vor der Einheit und 4 danach), ist die nicht durch das Modell erklärte Streuung im Osten mit 7 Prozentpunkten deutlich größer. Das ist kein Widerspruch zu dem höheren Bestimmtheitsmaß, sondern eine Folge der generell höheren Arbeitslosenquoten im Osten Deutschlands. Schließlich bedingen höhere Quoten im allgemeinen gleichzeitig eine größere Abweichung der gruppenspezifischen Quoten in den einzelnen Monaten von ihrem Erwartungswert. Diese Überlegung gilt zwar zunächst einmal nur für die Störgrößen, aber da die Störgrößen bei einem Wert für ρ in der Nähe von 0,25 zu drei Vierteln aus den Innovationen bestehen, überträgt sich die Argumentation auch auf diese.

Die Tabellen 9 bis 11 zeigen die Koeffizientenschätzungen für die Regressionen auf den drei Teildatensätzen. In der Referenzgruppe (Männer von 25 bis 49 Jahren mit mittlerem Bildungsabschluss im Monat Juni) ist die erwartete Arbeitslosenquote $\tilde{\alpha}$ in Westdeutschland mit 3,07 vor bzw. 3,72 nach der Einheit wesentlich niedriger als in Ostdeutschland mit 11,86. Bei den Frauen liegt der Wert in der gleichen demographischen Gruppe etwas höher, in Westdeutschland um 1,46 vor der Einheit und 0,22 nach der Einheit und in Ostdeutschland um 7,68. Die Abweichungen für die anderen demographischen Gruppen entsprechen im wesentlichen den Erwartungen. Für die Gruppe der gering Qualifizierten ist die Arbeitslosenquote höher als in der Referenzgruppe, während Personen mit einem hohen Bildungsabschluss seltener erwerbslos

sind. Die Parameter für die Gruppe der jungen Hochqualifizierten weisen für die Zeiträume vor bzw. nach der Wiedervereinigung unterschiedliche Vorzeichen auf (-1,82 vor der Einheit, 2,43 danach); beide Werte sind jedoch zum üblichen Signifikanzniveau von 5% nicht signifikant. Die Aussagekraft dieser Werte wird außerdem dadurch beeinträchtigt, dass die Gruppe sehr dünn besetzt ist (vgl. Unterkapitel 7.1). Auffällig ist die im Zeitraum nach der Einheit mit 13,34 (3,72 + 9,62) im Vergleich zum Zeitraum vor der Einheit mit 7,35 wesentlich höhere Arbeitslosenquote für die Gruppe der alten gering Qualifizierten. Hierbei ist allerdings zu berücksichtigen, dass der Anteil der gering Qualifizierten in dieser Altersgruppe von etwa einem Drittel vor der Einheit auf weniger als ein Viertel nach der Einheit gesunken ist (vgl. Tabelle 6, (a) und (b)). Allerdings sind die erwarteten Arbeitslosenquoten in der Gruppe der 50-64jährigen auch bei höheren Bildungsabschlüssen nach der Einheit wesentlich höher als zuvor, wenn auch auf niedrigerem Niveau (7,23 statt 3,19 bei mittlerem Bildungsabschluss, 3,0 statt 1,18 bei hohem Bildungsabschluss). Eine mögliche Erklärung für die gestiegenen Arbeitslosenquoten der älteren Personen könnte in den veränderten Rahmenbedingungen liegen (längere Bezugsdauer für Arbeitslosengeld, Altersteilzeit, generell erleichteter Übergang von der Arbeitslosigkeit in die Frührente).

Für westdeutsche Frauen ist der Einfluss der demographischen Gruppe weniger stark ausgeprägt als bei westdeutschen Männern; die Abweichungen $\tilde{\delta}_i$ von der Struktur der Männer $\tilde{\beta}_i$ weisen vielfach entgegengesetzte Vorzeichen auf und relativieren so die Abweichungen von der Referenzgruppe. Die Vorzeichen der $\tilde{\delta}_i$ bei niedrigem Bildungsabschluss sind beispielsweise negativ, so dass die Arbeitslosenquoten für gering qualifizierte Frauen im Vergleich zur Referenzgruppe weniger stark steigen als bei Männern.

Die geschätzten Monatseffekte entsprechen den Erwartungen: Insbesondere in den Wintermonaten sind die Arbeitslosenquoten höher als im Referenzmonat Juni. Diese Monatseffekte sind vor der Einheit stärker ausgeprägt als danach.

Tabelle 9: Regressionsergebnisse für das alte Bundesgebiet vor der Einheit

(asymptotische t -Statistiken in Klammern)

Absolutglied 3,07 Abweichung 1,46

(27,87) für Frauen (12,88)

Abweichung in den demographischen Gruppen

Alter	gesamt			Abweichung Frauen		
	Bildungsabschluss			Bildungsabschluss		
	niedrig	mittel	hoch	niedrig	mittel	hoch
jung	5,41	4,76	-1,82	-1,89	-2,17	0,85
(16-24)	(29,39)	(31,74)	(-1,20)	(-1,82)	(-9,58)	(0,43)
mittel	7,46	—	-1,55	-6,65	—	0,79
(25-49)	(39,01)		(-10,31)	(-24,25)		(2,78)
alt	4,28	0,12	-1,89	-1,29	1,09	2,92
(50-64)	(16,43)	(0,96)	(-7,78)	(-3,67)	(4,41)	(5,95)

Monatseffekte

Januar	Februar	März	April	Mai	Juni
1,18	1,11	0,74	0,35	0,05	—
(10,40)	(9,94)	(6,58)	(3,09)	(0,48)	
Juli	August	September	Oktober	November	Dezember
0,15	0,21	0,17	0,16	0,37	0,69
(1,32)	(1,85)	(1,47)	(1,34)	(3,19)	(5,95)

Konjunktoreffekte

Alter	Männer			Frauen		
	Bildungsabschluss			Bildungsabschluss		
	niedrig	mittel	hoch	niedrig	mittel	hoch
jung	-1,04	-1,35	0,42	-1,56	-0,99	0,51
(16-24)	(-14,18)	(-21,42)	(0,60)	(-19,23)	(-15,23)	(0,77)
mittel	-0,38	-0,30	-0,20	-0,42	-0,49	-0,43
(25-49)	(-4,75)	(-8,96)	(-3,35)	(-5,52)	(-11,27)	(-4,63)
alt	0,20	0,12	0,10	-0,20	0,44	-1,23
(50-64)	(1,89)	(2,61)	(1,02)	(-2,15)	(5,39)	(-6,92)

Tabelle 10: Regressionsergebnisse für das alte Bundesgebiet nach der Einheit

(asymptotische t -Statistiken in Klammern)

Absolutglied 3,72 Abweichung 0,22

(30,52) für Frauen (2,14)

Abweichung in den demographischen Gruppen

Alter	gesamt			Abweichung Frauen		
	Bildungsabschluss			Bildungsabschluss		
	niedrig	mittel	hoch	niedrig	mittel	hoch
jung	5,44	3,50	2,43	-0,15	-1,93	-4,54
(16-24)	(24,82)	(18,45)	(1,93)	(-4,54)	(-7,11)	(-2,79)
mittel	7,76	—	-1,64	-3,38	—	1,67
(25-49)	(42,62)		(-13,02)	(-12,79)		(7,64)
alt	9,62	3,51	-0,72	-3,21	2,87	-1,26
(50-64)	(34,05)	(29,86)	(-3,95)	(-8,92)	(14,48)	(-3,09)

Monatseffekte

Januar	Februar	März	April	Mai	Juni
0,36	0,39	0,22	0,03	-0,04	—
(2,41)	(2,68)	(1,50)	(0,26)	(-0,31)	
Juli	August	September	Oktober	November	Dezember
0,19	0,08	-0,10	0,14	0,23	0,15
(1,34)	(0,54)	(-0,70)	(0,99)	(1,62)	(1,04)

Konjunkturreffekte

Alter	Männer			Frauen		
	Bildungsabschluss			Bildungsabschluss		
	niedrig	mittel	hoch	niedrig	mittel	hoch
jung	-0,86	-0,72	0,13	-0,31	-0,59	-0,20
(16-24)	(-13,02)	(-11,98)	(0,22)	(-4,36)	(-9,78)	(-0,51)
mittel	-0,87	-0,36	-0,27	-0,62	-0,23	-0,17
(25-49)	(-14,18)	(-14,44)	(-6,56)	(-10,03)	(-7,57)	(-2,75)
alt	-0,40	-0,46	-0,37	-0,61	-0,76	-0,06
(50-64)	(-4,43)	(-12,97)	(-5,85)	(-8,68)	(-14,66)	(-0,45)

Tabelle 11: Regressionsergebnisse für das neue Bundesgebiet

(asymptotische t -Statistiken in Klammern)

Absolutglied 11,86 Abweichung 7,68

(49,35) für Frauen (34,30)

Abweichung in den demographischen Gruppen

Alter	gesamt			Abweichung Frauen		
	Bildungsabschluss			Bildungsabschluss		
	niedrig	mittel	hoch	niedrig	mittel	hoch
jung	-5,61	5,32	18,26	-3,53	-11,61	-27,95
(16-24)	(-11,56)	(11,78)	(4,32)	(-4,52)	(-18,59)	(-6,20)
mittel	8,65	—	-9,52	2,48	—	-4,97
(25-49)	(14,65)		(-31,19)	(2,85)		(-12,05)
alt	19,63	10,47	5,94	0,63	0,24	-11,34
(50-64)	(21,32)	(34,37)	(16,67)	(0,54)	(0,53)	(-20,80)

Monatseffekte

Januar	Februar	März	April	Mai	Juni
1,91	1,87	1,23	0,53	0,13	—
(7,10)	(6,99)	(4,60)	(1,97)	(0,48)	
Juli	August	September	Oktober	November	Dezember
0,61	0,62	0,50	0,48	0,49	1,02
(2,29)	(2,30)	(1,87)	(1,80)	(1,82)	(3,78)

Konjunkturreffekte

Alter	Männer			Frauen		
	Bildungsabschluss			Bildungsabschluss		
	niedrig	mittel	hoch	niedrig	mittel	hoch
jung	0,00	-0,85	-3,06	-0,06	0,28	-0,47
(16-24)	(0,07)	(-11,81)	(-2,80)	(-0,53)	(4,07)	(-1,74)
mittel	1,45	-0,59	0,23	2,16	0,15	-0,00
(25-49)	(9,06)	(-21,46)	(4,83)	(15,51)	(4,92)	(-0,06)
alt	1,36	-1,45	-1,01	2,30	-0,54	-0,80
(50-64)	(6,48)	(-29,57)	(-16,79)	(18,16)	(-10,13)	(-11,19)

Die Konjunkturreffekte schließlich geben an, wie sich die Arbeitslosenquoten in den demographischen Gruppen ändern, wenn das Wirtschaftswachstum im Vorjahr um ein Prozent höher liegt. Dementsprechend sind die Koeffizienten für die meisten Gruppen negativ - wenn die Wirtschaft im Vorjahr stärker gewachsen ist, sind die Arbeitslosenquoten niedriger. Der Effekt ist bei den unter 25jährigen tendenziell stärker ausgeprägt als bei den anderen Altersklassen. Die mitunter sogar positiven Effekte für junge Hochqualifizierte sind nicht signifikant und dürften auf die geringen Stichprobenumfänge dieser Gruppe zurückzuführen sein. Überraschend sind hingegen die positiven Effekte bei Männern in der höchsten Altersgruppe vor der Einheit.

Tabelle 11 zeigt die Ergebnisse für den ostdeutschen Datensatz. Hier ist zunächst einmal festzustellen, dass die Arbeitslosenquoten generell höher sind als im Westen. In der Referenzgruppe der Männer ergibt sich beispielsweise ein Effekt von 11,86 und für die entsprechende Gruppe der Frauen beträgt der Effekt sogar 19,54. Die Abweichungen der anderen demographischen Gruppen von der Referenzgruppe sind weniger systematisch als in Westdeutschland. Auffällig sind die niedrigen Werte für hoch Qualifizierte der mittleren Altersgruppe (2,34 bei den Männern und 5,05 bei den Frauen) sowie die hohen Werte für gering Qualifizierte der höchsten Altersgruppe (31,49 bei den Männern und 39,8 bei den Frauen). Die mit 30,12 extrem hohe Arbeitslosenquote für junge hoch qualifizierte Männer dürfte - genau wie im Westen - auf die dünne Besetzung dieser Klasse zurückzuführen sein; für Frauen der gleichen Gruppe liegt der Wert bei 9,85. Die Monateffekte zeigen tendenziell die gleiche saisonale Struktur wie im Westen, sind allerdings etwas stärker ausgeprägt.

Die oben beschriebenen Resultate ergeben sich unter Einbeziehung der räumlichen Korrelationen der Störgrößen. Im Hinblick auf den Datensatz ist diese Modellierung plausibel, denn die häufig getroffene Annahme einer skalaren Kovarianzmatrix der Störgrößen lässt sich aus sachlogischen Überlegungen kaum rechtfertigen. Auf der anderen Seite bedeutet die feinere

Tabelle 12: Einfache KQ-Regression für den Zeitraum vor der Einheit

Monatseffekte					
Januar	Februar	März	April	Mai	Juni
0,64	0,50	0,63	0,24	-0,00	—
(1,79)	(1,39)	(1,76)	(0,67)	(-0,01)	
Juli	August	September	Oktober	November	Dezember
0,80	0,43	-0,02	0,05	-0,09	0,16
(2,14)	(1,17)	(-0,06)	(0,14)	(-0,25)	(0,42)

Modellierung im Vergleich mit der gewöhnlichen KQ-Methode einen erhöhten Aufwand. Dies führt zu der Frage, ob sich der größere Aufwand überhaupt lohnt: Unterscheiden sich die oben beschriebenen Ergebnisse stark von denen einer gewöhnlichen KQ-Regression oder sind die Unterschiede gering, so dass man sich den Aufwand hätte sparen können? Die Antwort fällt positiv aus: Ja, der Aufwand lohnt sich. Bei einer gewöhnlichen KQ-Regression ergeben sich nicht nur wesentlich kleinere korrigierte Bestimmtheitsmaße (0,414 statt 0,753 für Westdeutschland vor der Einheit, 0,363 statt 0,785 für Westdeutschland nach der Einheit sowie 0,775 statt 0,914 für den Osten). Auch die Parametertests für die einzelnen Komponenten von β kommen zu unterschiedlichen Entscheidungen. Während die Schätzwerte selbst ähnlich sind, unterscheiden sich deren geschätzte Standardabweichungen stark voneinander: Die geschätzten Schätzfehler sind bei der gewöhnlichen KQ-Regression größer als unter Berücksichtigung der räumlich korrelierten Struktur. Dadurch sind zahlreiche Effekte nicht mehr signifikant. So sind etwa von den 18 t -Statistiken der Konjunkturreffekte für den Zeitraum vor der Einheit 14 betragsmäßig größer als 2 (vgl. Tabelle 9) und somit zum 5%-Niveau signifikant; bei einer gewöhnlichen KQ-Regression ist das nur noch für 6 dieser Effekte der Fall.

Starke Unterschiede gibt es auch bei den Monatseffekten: Tabelle 9 zeigt,

dass bei Berücksichtigung der räumlichen Korrelationen die Arbeitslosenquoten vor allem in den Wintermonaten signifikant höher sind als im Referenzmonat Juni. Zum Vergleich stellt Tabelle 12 die entsprechenden Ergebnisse bei einer gewöhnlichen KQ-Regression dar. Hier ist nur der Julieffekt signifikant (positiv); für die Effekte der Wintermonate ist die Null in den jeweiligen 95%-Konfidenzintervallen enthalten.

Insgesamt lohnt sich die Modellierung der Korrelationsstruktur, weil sich vor allem die Ergebnisse der Signifikanztests von denen bei einer herkömmlichen KQ-Regression deutlich unterscheiden. Aber auch die Schätzwerte selber - wie zum Beispiel bei den Monateffekten - stimmen eher mit substanzwissenschaftlichen Überlegungen überein.

8 Zusammenfassung und Fazit

In Regressionsmodellen wird häufig vereinfachend angenommen, dass die Störgrößen alle die gleiche Varianz besitzen und zudem untereinander unkorreliert sind. Insbesondere die zweite dieser Annahmen ist bei vielen Datensätzen jedoch unrealistisch, da bestimmte Korrelationsstrukturen plausibel erscheinen. Es gibt viele Möglichkeiten, solche Strukturen zu modellieren. Diese Arbeit beschäftigt sich mit einer davon: mit dem Modell räumlich korrelierter Störgrößen.

Eine Schätzung der Modellparameter nach Maximum-Likelihood-Methode erweist sich insbesondere bei großen Datensätzen als rechenaufwändig und numerisch instabil. Als Alternative bietet sich ein in der Literatur häufig verwendeter Schätzer nach generalisierter Momentenmethode an. Dieser Momentenschätzer setzt Erwartungswerte quadratischer Formen des Störgrößenvektors mit den entsprechenden empirischen Momenten gleich. Da die Störgrößen selbst unbeobachtbar sind, dienen die beobachtbaren Regressionsresiduen zur Berechnung der empirischen Momente. Asymptotisch sind die Unterschiede zwischen Störgrößen und Regressionsresiduen zu vernachlässigen, aber für jeden endlichen Stichprobenumfang stimmen sie nicht exakt überein. Diese Überlegung führt auf eine naheliegende Modifikation des Schätzers nach generalisierter Momentenmethode: Wenn die empirischen Momente aus den Regressionsresiduen berechnet werden müssen, dann sollten auch für die theoretischen Momente die Regressionsresiduen an die Stelle der Störgrößen treten. Der aus der Modifikation resultierende Schätzer ist ebenfalls leicht zu berechnen und konsistent.

Die Modifikation berücksichtigt den Unterschied zwischen Störgrößen und Regressionsresiduen. Die Anwendung dieser Idee ist nicht auf den Schätzer nach verallgemeinerter Momentenmethode beschränkt: Sie lässt sich auf den Maximum-Likelihood-Schätzer übertragen. Der Vorteil liegt hier im wesentlichen in einer Rechenerleichterung. Auch auf Residuen aus einer verallgemeinerten KQ-Regression oder einer nichtlinearen Regression ist die Modifi-

kation anwendbar.

Eine Simulationsstudie verdeutlicht, wie sich der Schätzer nach verallgemeinerter Momentenmethode durch die Modifikation verbessert: Der mittlere quadratische Fehler bei einer Schätzung der Modellparameter geht deutlich zurück. Diese Verbesserung beruht zu großen Teilen auf einer um etwa drei Viertel reduzierten Verzerrung der Schätzer. Da die Schätzer auch ein Bestandteil der geschätzten Kovarianzmatrix des KQ-Schätzers für den Vektor der Regressionskoeffizienten sind, bewirkt die Modifikation einen weiteren angenehmen Effekt: Die Signifikanztests für die Regressionskoeffizienten überschreiten die vorgegebene Wahrscheinlichkeit für einen Fehler erster Art deutlich weniger als ohne die Modifikation. Dieser Effekt tritt auch noch bei großen Stichprobenumfängen auf.

Gewissermaßen als Nebenprodukt offenbart die Simulation auch noch weitere interessante Zusammenhänge. So unterscheiden sich die tatsächlichen Niveaus der Signifikanztests für verschiedene Regressionskoeffizienten deutlich voneinander. Auch die Schätzung der Modellparameter variiert in Abhängigkeit von den wahren Werten. Diese in der Simulation entdeckten Zusammenhänge können mit Methoden der Matrixdifferentiation auch analytisch begründet werden.

Klassische Anwendungsgebiete für räumliche Korrelationen sind beispielsweise Bergbau, Ökologie oder Epidemiologie. Schon die Bezeichnung „räumliche“ Korrelationen deutet an, für welche Art von Daten das Modell ursprünglich entwickelt wurde. Die Merkmalsträger sind häufig verschiedene Gebiete, die in einer bestimmten räumlichen Beziehung zueinander stehen. Die Einsatzmöglichkeiten des Modells sind jedoch nicht auf diese klassischen Anwendungen beschränkt. Auch verschiedene soziale Gruppen können als räumlich korreliert modelliert werden. Dazu wird festgelegt, welche Gruppen in direkter Beziehung zueinander stehen. Die Anwendung in dieser Arbeit analysiert aus dem sozio-ökonomischen Panel berechnete Arbeitslosenquoten in verschiedenen demographischen Gruppen der Bevölkerung. Das un-

terstellte Regressionsmodell berücksichtigt nicht nur die räumlichen Korrelationen, sondern auch die unterschiedlichen Varianzen der einzelnen Arbeitslosenquoten. Die Ergebnisse stimmen besser mit substanzwissenschaftlichen Überlegungen überein als die Ergebnisse einer einfachen KQ-Regression. Der durch die genauere Abbildung der Kovarianzstruktur entstehende zusätzliche Aufwand ist somit gerechtfertigt.

Die in dieser Arbeit vorgeschlagene Modifikation eines generalisierten Momentenmethodenschätzers erreicht eine Verbesserung des ursprünglichen Schätzers, indem die Unterschiede zwischen unbeobachtbaren Störgrößen und beobachtbaren Regressionsresiduen berücksichtigt werden. Darüber hinaus sind (zumindest) zwei weitere Ansatzpunkte zur Verbesserung des Schätzers denkbar.

Der erste Punkt bezieht sich auf die verwendeten Momente. Um drei verschiedene theoretische Momente zu erhalten, wird der Vektor der Residuen teilweise von links mit der Verwandtschaftsmatrix multipliziert. Das ist naheliegend, weil die Verwandtschaftsmatrix das passende Format besitzt. Es ist aber durchaus möglich, dass die Multiplikation mit einer anderen Matrix passenden Formats bessere Resultate liefert. Interessant ist dann die Frage, für welche Matrix die besten Resultate erreicht werden können.

Ein zweiter Ansatzpunkt für eine weitere Verbesserung ist die zu minimierende Zielfunktion. Sie enthält die ungewichteten quadrierten Abstände zwischen den drei theoretischen und den drei entsprechenden empirischen Momenten. Die drei Momente scheinen auf den ersten Blick gleichberechtigt. Nichtsdestotrotz könnte eine geeignete Gewichtung der Abstände zu einer weiteren Verbesserung des Schätzers beitragen.

Tabellenverzeichnis

1	Kennzahlen der Schätzer für $n = 20$	49
2	Kennzahlen der Schätzer für $n = 60$	50
3	Kennzahlen der Schätzer für $n = 100$	51
4	Kennzahlen der Schätzer für $n = 400$	52
5	Überblick über den Datensatz	63
6	Verteilung der Stichprobe auf die demographischen Gruppen .	65
7	Anzahl direkter Nachbarn	69
8	Übersicht über die Regressionen	74
9	Regressionsergebnisse für das alte Bundesgebiet vor der Einheit	76
10	Regressionsergebnisse für das alte Bundesgebiet nach der Einheit	77
11	Regressionsergebnisse für das neue Bundesgebiet	78
12	Einfache KQ-Regression für den Zeitraum vor der Einheit . .	80

Abbildungsverzeichnis

1	Beispiel für Verwandtschaftsmatrix	6
2	Einfluss des räumlichen Autokorrelationsparameters	57
3	Reales jährliches Wachstum des Bruttoinlandsprodukts	72

Literatur

- L. Anselin, R. Bongiovanni, und J. Lowenberg-DeBoer. A spatial econometric approach to the economics of site-specific nitrogen management in corn production. *American Journal of Agricultural Economics*, 86(3):675–687, 2004.
- K. P. Bell und N. E. Bockstael. Applying the generalized-moments estimation approach to spatial problems involving microlevel data. *The Review of Economics and Statistics*, 82(1):72–82, 2000.
- A.D. Cliff und J.K. Ord. *Spatial Autocorrelation*. Pion, London, 1973.
- V. Druska und W. C. Horrace. Generalized moments estimation for spatial panel data: Indonesian rice farming. *American Journal of Agricultural Economics*, 86(1):185–198, 2004.
- R. Ezcurra, P. Pascual, und M. Rapun. The spatial distribution of welfare in the European Union. *Tijdschrift voor Economische en Sociale Geografie*, 97(4):331–342, 2006.
- W. Franz. *Arbeitsmarktökonomik*. Springer, Berlin, Heidelberg, New York, 2003.
- W. A. Fuller. *Introduction to Statistical Time Series*. John Wiley & Sons, New York, 1996.
- T. A. Garrett und D. C. Wheelock. Why did income growth vary across states during the great depression? *Economic History Association*, 66(2): 456–466, 2006.
- R. I. Jennrich. Asymptotic properties of non-linear least squares estimators. *Annals of Mathematical Statistics*, 40(2):633–643, 1969.

- H.H. Kelejian und I.R. Prucha. A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, 40:509–533, 1999.
- H.H. Kelejian und I.R. Prucha. Estimation of simultaneous systems of spatially interrelated cross sectional equations. *Journal of Econometrics*, 118: 27–50, 2004.
- J. Kluve, S. Schaffner, und C.M. Schmidt. Labor force status dynamics in the German labor market: Individual heterogeneity and cyclical sensitivity. *Technical Report, SFB 475, University of Dortmund*, 20:1–50, 2006.
- W. Krämer. The robustness of the F -test to spatial autocorrelation among regression disturbances. *Statistica*, 3:435–440, 2003.
- W. Krämer. Finite sample power of cliff-ord-type tests for spatial disturbance correlation in linear regression. *Journal of Statistical Planning and Inference*, 128:489–496, 2005.
- W. Krämer und C. Donninger. Spatial autocorrelation among errors and the relative efficiency of OLS in the linear regression model. *Journal of the American Statistical Association*, 82:577–579, 1987.
- J. R. Magnus und H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons, New York, 1988.
- E. Malinvaud. *Statistical Methods of Econometrics*. North Holland, Amsterdam, 1970.
- A. W. Marshall und I. Olkin. *Inequalities: Theory of Majorization and Its Applications*. Academic Press, San Diego, 1979.

- W. Oberhofer und J. Kmenta. A general procedure for obtaining maximum likelihood estimates in generalized regression models. *Econometrica*, 42(3): 579–590, 1974.
- B. M. Pötscher und I. R. Prucha. Basic structure of the asymptotic theory in dynamic nonlinear econometric models, part i: Consistency and approximation concepts. *Econometric Reviews*, 10(2):125–216, 1991.
- W. Schlenker, W. M. Hanemann, und A. C. Fisher. The impact of global warming on U.S. agriculture: An econometric analysis of optimal growing conditions. *Review of Economics and Statistics*, 88(1):113–125, 2006.
- K. Schmidt und G. Trenkler. *Einführung in die Moderne Matrix-Algebra - Mit Anwendungen in der Statistik*. Springer, Berlin, Heidelberg, 2006.
- G.A.F. Seber und A.J. Lee. *Linear Regression Analysis*. John Wiley & Sons, Hoboken, New Jersey, 2003.
- P. G. Tingvall. The dynamics of European industrial structure. *Review of World Economics*, 140(4):665–687, 2004.
- H. White. *Asymptotic Theory for Econometricians*. Academic Press, Orlando, 1984.
- P. Whittle. On stationary processes in the plane. *Biometrika*, 41:434–449, 1954.