

### Ein Messwiederholungsmodell mit zufälligen Koeffizienten in der Regionalforschung: Anwendung am Beispiel räumlicher Determinanten der Fluktuation zwischen regionalen Arbeitsmärkten in Westdeutschland 1984-1997

Windzio, Michael

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

#### Empfohlene Zitierung / Suggested Citation:

Windzio, M. (2004). Ein Messwiederholungsmodell mit zufälligen Koeffizienten in der Regionalforschung: Anwendung am Beispiel räumlicher Determinanten der Fluktuation zwischen regionalen Arbeitsmärkten in Westdeutschland 1984-1997. *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, 55, 6-33. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-198715>

#### Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

#### Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

## **Ein Messwiederholungsmodell mit zufälligen Koeffizienten in der Regionalforschung**

### **Anwendung am Beispiel räumlicher Determinanten der Fluktuation zwischen regionalen Arbeitsmärkten in Westdeutschland 1984-1997**

von Michael Windzio <sup>1</sup>

#### ***Zusammenfassung***

*In dem Aufsatz wird das Potential eines Modells der Längsschnittanalyse dargestellt und dessen Anwendung am Beispiel einer Aggregatdatenanalyse der Mobilität von Beschäftigten zwischen regionalen Arbeitsmärkten demonstriert. Es handelt sich um ein auf kurzen Zeitreihen von interregionalen Mobilitätsquoten basierendes Messwiederholungsmodell, bei dem die Mobilität sowohl von direkt gemessenen erklärenden Variablen – wie beispielsweise der regionalen Arbeitslosenquote – als auch von latenten zeitlichen Trends abhängig sein kann. Dies ermöglichen zufällige Koeffizienten, mit deren Hilfe für jede Region unter Kontrolle der manifesten Einflussfaktoren spezifische Entwicklungen erfasst werden können. Schließlich wird auf die Umsetzung des Modells in der SAS mixed Prozedur hingewiesen.*

#### ***Abstract***

*This paper deals with the random coefficient model in longitudinal analysis. Its potential benefits are shown by referring on an example from labour market research, investigating mobility between local labour markets by using short time series of aggregate data. In the model, the rate of interregional mobility depends on directly measured characteristics like regional unemployment rates. Furthermore, mobility is influenced by corresponding unobserved factors which are introduced into the model as random coefficients. By these random coefficients historical trends can be estimated conditionally on directly measured characteristics separately for each*

---

1 Dr. **Michael Windzio** ist als Wissenschaftlicher Mitarbeiter am Kriminologischen Forschungsinstitut Niedersachsen, Lützerodestraße 9, 30161 Hannover, im Bereich der quantitativen Methoden tätig.

region. Finally, the implementation of this model in the SAS mixed procedure is shown.

## 1 Einleitung<sup>2</sup>

In der vergleichenden ökonomischen und politikwissenschaftlichen Forschung liefern auf Zeitreihen basierende Regressionsmodelle die Möglichkeit, Einflussfaktoren auf die abhängige Variable veränderlich über die Zeit zu schätzen. Von jeder Untersuchungseinheit – das können Nationen, Betriebe, Wahlbezirke oder auch Regionen sein – geht dabei nicht nur eine Messung in die Schätzung ein, sondern ebenso viele Messungen wie die Zeitreihen Beobachtungspunkte umfassen. Weil diese Messungen nicht unabhängig voneinander sind und darum gegen Annahmen des klassischen OLS Regressionsmodells verstoßen, wurden einige Techniken zur Lösung der damit verbundenen Probleme entwickelt (**Beck** and **Katz** 1995, vgl. **Kittel** and **Winner** 2002). In jüngerer Zeit wurde in der Panelanalyse das überaus flexible Modell mit zufälligen Koeffizienten (Random Coefficient Model) in Softwareprogrammen wie SAS, STATA oder MLwiN implementiert, bei dem die geschätzten Effekte der zeitveränderlichen erklärenden Variablen zwischen den jeweiligen Untersuchungseinheiten variieren können. Dieser Ansatz wurde bereits in den 1970er Jahren insbesondere von **Swamy** (1970) für die Analyse komplexer Datenstrukturen mit gemischten Verteilungen entwickelt (**Littell** u.a. 1996) und auf die Analyse von Zeitreihen übertragen (**Swamy** 1970, **Western** 1998). In diesem Zusammenhang wird es als „Time Series Cross Section Analysis“ (TSCS) bezeichnet.

Es existieren unterschiedliche Bezeichnungen für dieses Modell, die von der disziplinären Herkunft der Forschenden, vor allem aber von dem Verhältnis der Anzahl der Messzeitpunkte pro Objekt zu der Anzahl der Objekte selbst abhängig sind: so ist die *TSCS Analyse* eine Messwiederholungsanalyse, bei der man Prozesse einer eher geringen Anzahl von Objekten in der Regel über einen längeren Zeitraum verfolgt. Die asymptotischen Eigenschaften der Schätzung werden für eine Situation unterstellt, in der die Anzahl der *Zeitpunkte* theoretisch gegen unendlich konvergiert (**Greene** 2000: 593). Dagegen basiert die *Panelanalyse* auf der Annahme, dass dies für die Zahl der *Objekte* gilt, nicht jedoch für die Zahl der Panelwellen. Kommen zufällige Regressionskonstanten oder gar zufällige Koeffizienten hinzu, hat

---

2 Die Analysen dieser Arbeit entstanden am Institut für empirische und angewandte Soziologie der Universität Bremen (EMPAS) sowie im Kontext des Projekts C4 des Sonderforschungsbereichs 597. Für überaus wertvolle Hinweise und Kommentare danke ich einem Gutachter/ einer Gutachterin der ZA-Information und den Teilnehmern der Session „Time Series Analysis“ der Sixth International ISA Conference on Logic and Methodology in Amsterdam (August 2004).

sich in der Soziologie und Psychologie die Bezeichnung „multilevel model for change“ (*Singer* and *Willet* 2003: 45) etabliert.

Das Messwiederholungsmodell mit zufälligen Koeffizienten soll im Folgenden dargestellt und dessen Vorteile für die Modellierungen zeitlicher Veränderungen hervorgehoben werden. Als Beispiel dient eine Analyse der überregionalen Arbeitsmarktmobilität.

## 2 Inhaltliche Fragestellung

Die (In-)Flexibilität von Arbeitsmärkten ist ein breit diskutiertes Thema in der öffentlichen Debatte. Im Mittelpunkt der ökonomischen und soziologischen Forschung stehen insbesondere arbeitsrechtliche Fragen und betriebliche Faktoren, durch die eine Anpassung der Beschäftigung an ökonomische Erfordernisse gefördert oder erschwert wird (*Struck* und *Köhler* 2004). Aus geographischer Sicht liegt es hingegen nahe, auch nach räumlichen Differenzierungen zu fragen, die hinsichtlich der Flexibilität von Beschäftigungsverhältnissen zu vermuten sind. Welche Regionen Westdeutschlands weisen vergleichsweise hohe oder geringe überregionale Ein- und Austrittsquoten von Arbeitskräften auf? Und wie entwickelte sich die überregionale Mobilität über die Zeit?

In der vorliegenden Arbeit wird zunächst die inhaltliche Fragestellung näher ausgeführt, dann das Messwiederholungsmodell mit zufälligen Effekten vorgestellt sowie ein deskriptiver Überblick über die räumliche Mobilität in Westdeutschland von 1984 bis 1997 und deren zeitlichen Verlauf geliefert. Schließlich wird mit Hilfe des Messwiederholungsmodells (vgl. Abschnitt 4) geprüft, durch welche regionalen Merkmale die Ein- und Austrittsquoten auf regionalen Arbeitsmärkten im Zeitverlauf zu erklären sind.

## 3 Einflussfaktoren auf die regionale Mobilität: Hypothesen

Das aus den 1960er Jahren stammende und in der ökonomischen Migrationsforschung mittlerweile klassische so genannte *Lowry*-Modell beinhaltet eine ökonomische Erweiterung des auf Masse und Distanz basierenden Gravitationsmodells um die regionale Arbeitslosenquote und die mittleren Löhne im Manufakturbereich (*Boyle, Halfacree* and *Robinson* 1998: 90f):

$$\ln(\hat{M}_{ij}) = a_0 + a_1 \ln(u_i) + a_2 \ln(u_j) + a_3 \ln(w_i) + a_4 \ln(w_j) + a_5 \ln(L_i) + a_6 \ln(L_j) + b \ln(d_{ij})$$

In diesem Modell ist  $\hat{M}_{ij}$  die Rate der Migration zwischen den Regionen  $i$  und  $j$ ,  $a_0$  ist die Regressionskonstante; als Äquivalente zum Masse-Term der Gravitations-theorie sind  $L_i$  und  $L_j$  die jeweilige Anzahl der Beschäftigten in den Regionen  $i$  und  $j$ ,  $d_{ij}$  ist die Distanz zwischen den Regionen  $i$  und  $j$ ;  $u_i$  bzw.  $u_j$  sind die regionalen Arbeitslosenquoten und  $w_i$  bzw.  $w_j$  die mittleren regionalen Löhne. Obwohl es sich um ein Aggregatmodell handelt, basiert die ökonomische Erklärung der Effekte auf der Annahme nutzenmaximierender Akteure, die von Regionen mit hohen mittleren Löhnen angezogen werden, Regionen mit hoher Arbeitslosigkeit dagegen meiden und im Zuge der Migration im Aggregat eine Tendenz zum regionalen Gleichgewicht fördern.

Im Gegensatz zum **Lowry**-Modell ist in der vorliegenden Arbeit die Frage der Fluktuation auf regionalen Arbeitsmärkten nicht auf den Austausch zwischen zwei Regionen  $i$  und  $j$  bezogen, sondern auf die Anteile der Ein- und Austritte von Arbeitskräften in einer Region an allen dort während einer Zeitperiode Beschäftigten. In Anlehnung an die Darstellung von **Boyle, Halfacree** und **Robinson** (1998: 45) werden also *in-migration rates* und *out-migration rates* getrennt untersucht. Neben den ökonomischen Faktoren des mittleren regionalen Einkommens und der Arbeitslosenquote sowie dem Effekt der Masse sind Einflüsse weiterer regionaler Merkmale zu vermuten, die in den Regressionsmodellen zumindest zu kontrollieren sind (vgl. unten Abschnitt 5). In Verbindung mit weiteren Kontrollvariablen werden im Folgenden die aus dem **Lowry**-Modell abgeleiteten Hypothesen über nutzenmaximierende Akteure, die Regionen mit hohen mittleren Gehältern und niedrigen Arbeitslosenquoten suchen bzw. Regionen mit geringen mittleren Gehältern und hohen Arbeitslosenquoten meiden, geprüft. Dies geschieht auf Basis der folgenden Modelle:

Eintritte in Region  $j$ :

$$E_j = \beta_0 + \beta_1(T) + \beta_2(T^2) + \beta_3 \ln(N) + \beta_4(DI) + \beta_5(Alo) + \beta_6(Arme) + \beta_7(D\_Bahn) \\ + \beta_8(Verd) + \beta_9(Bev\_w) + \beta_{10}(Ak) + \beta_{11}(Eink) + \beta_{12}(Einkv) + \beta_{13}(Ausl) + \beta_{14-21}(Kreistypen)$$

Austritte aus Region  $j$ :

$$A_j = \beta_0 + \beta_1(T) + \beta_2(T^2) + \beta_3 \ln(N) + \beta_4(DI) + \beta_5(Alo) + \beta_6(Arme) + \beta_7(D\_Bahn) \\ + \beta_8(Verd) + \beta_9(Bev\_w) + \beta_{10}(Ak) + \beta_{11}(Eink) + \beta_{12}(Einkv) + \beta_{13}(Ausl) + \beta_{14-21}(Kreistypen)$$

**Tabelle 1** Zusammenfassung der Hypothesen über die Einflussfaktoren auf die regionalen Ein- und Austritte

	Eintritte	Effekt	Austritte	Effekt
<i>Jahre seit 1984</i>	$\beta_1(T)$	*	$\beta_1(T)$	*
<i>Jahre<sup>2</sup> seit 1984</i>	$\beta_2(T^2)$	*	$\beta_2(T^2)$	*
<i>ln(Beschäftigtenzahl)</i>	$\beta_3 \ln(N)$	+	$\beta_3 \ln(N)$	–
<i>% Besch. wirtsch. Dienstleistungen</i>	$\beta_4(DI)$	+	$\beta_4(DI)$	+
<i>Arbeitslosenquote</i>	$\beta_5(Alo)$	–	$\beta_5(Alo)$	+
<i>% Arme (Bezieher von HLU)</i>	$\beta_6(Arme)$	–	$\beta_6(Arme)$	+
<i>Distanz Umschlagbahnhof (Pkw min.)</i>	$\beta_7(D\_Bahn)$	*	$\beta_7(D\_Bahn)$	*
<i>Distanz zu Verdichtungsraum</i>	$\beta_8(Verd)$	–	$\beta_8(Verd)$	–
<i>Bevölkerungswachstum 1987-97</i>	$\beta_9(Bev\_w)$	+	$\beta_9(Bev\_w)$	+
<i>% Akademiker</i>	$\beta_{10}(Ak)$	+	$\beta_{10}(Ak)$	+
<i>Einkommen</i>	$\beta_{11}(Eink)$	+	$\beta_{11}(Eink)$	–
<i>Einkommensvarianz</i>	$\beta_{12}(Einkv)$	*	$\beta_{12}(Einkv)$	*
<i>% Ausländer</i>	$\beta_{13}(Ausl)$	*	$\beta_{13}(Ausl)$	*

Anmerkung: + positiver Effekt, – negativer Effekt, \* reine Kontrollvariable

Interessanterweise ermittelte **Lowry** selbst einen nicht signifikanten negativen Effekt für die Arbeitslosenquote der Herkunftsregion, was entweder die Frage aufwirft, durch welche externen Faktoren die Gleichgewichtstendenz gestört ist oder in welcher besonderen Form die Arbeitslosenquote die Mobilität beeinflusst. In Tabelle 1 sind die Hypothesen sowie die Einflüsse der Kontrollvariablen zusammengefasst.

#### 4 Daten und Methoden

Das Institut für Arbeitsmarkt- und Berufsforschung (IAB) hat der Wissenschaft mit dem Regionalfile<sup>3</sup> eine 1% Stichprobe aller sozialversicherungspflichtigen (sv) Beschäftigten zwischen 1975 und 1997 zugänglich gemacht, in der unter anderem die Kreiskennziffer des jeweiligen Betriebes enthalten ist (**Haas** 2001: 131). Leider sind die Angaben über die Kreise nicht in allen Fällen exakt, da das IAB aus Gründen der Anonymisierung einige Kreise zusammenlegen musste. Es wurde angestrebt, möglichst Kreise identischer Typen zu aggregieren. Allerdings wurde gegenüber dem Kriterium identischer Typen dem der räumlichen Nachbarschaft Vorrang eingeräumt, so dass ein Teil der Kreise nicht korrekt durch den Kreistyp identifiziert ist (**Haas** 2001: 133, vgl. Tabelle A1 im Anhang).

3 Dieser Datensatz ist in der Zugangsgruppe C unter der Studiennummer 3348 am Zentralarchiv zugänglich.

Zwei abhängige Variablen wurden für die Messwiederholungsanalyse aus den Individualdaten des IAB aggregiert: erstens die Logits für *Eintritte* von Beschäftigten in einen regionalen Arbeitsmarkt und zweitens die Logits von *Austritten* von Beschäftigten aus einem regionalen Arbeitsmarkt. Erforderlich wurde die Logit-Transformation, wie sie unter anderem bei **Greene** (2000: 834f) dargestellt ist, weil eine Schätzung von Anteilswerten als abhängige Variable ähnliche Probleme aufweist, wie das so genannte lineare Wahrscheinlichkeitsmodell zur Schätzung von Ereignissen im binär codierten Zustandsraum (vgl. **Long** 1997: 35ff). Bei den Logits der Eintritte wurde der Anteil aller in einem Jahr in einen regionalen Arbeitsmarkt eintretenden Personen an allen in diesem Jahr dort Beschäftigten durch die Gegenwahrscheinlichkeit dividiert und das Resultat logarithmiert. In analoger Weise wurde mit den Austritten verfahren (vgl. **Milne** 1991: 144). Allerdings wurden bei den Eintritten die Berufseinsteiger, bei den Austritten die dauerhaften Aussteiger aus dem Erwerbssystem ausgeschlossen. Als Eintretende wurden also nur Personen betrachtet, die *zuvor*, und als Austretende hingegen nur Personen, die *anschließend* in einer anderen Region sv beschäftigt waren.

Für die Analyse der Modelle in den Gleichungen (1) und (2) erhalten wir bei 270 westdeutschen Regionen in den 14 Jahren von 1984-1997 insgesamt 3780 Messungen. Tabelle 2 zeigt die Mittelwerte und Standardabweichungen der Variablen.

**Tabelle 2** Mittelwerte und Standardabweichungen der Variablen

<b>Variable</b>	<b>Mean</b>	<b>stddev</b>
Logit(Eintritte)	-2,58361	0,35860
Logit(Austritte)	-2,57580	0,33690
T (Jahre seit 1984, metrisch)	6,5	4,03166
T <sup>2</sup> (Jahre quadriert)	58,5	54,3602
Ln(Beschäftigte in Region)	6,71522	0,62005
% Beschäftigte wirtschaftsnahen Dienstleistungen	27,4143	7,36815
Arbeitslosenquote	8,63031	3,39595
% Armut (HLU)	37,3452	19,2249
Distanz Umschlagbahnhof (min.)	28,7	23,9470
Distanz Verdichtungsraum (min.)	23,9815	22,4876
Δ Bevölkerung 1987-1997	10,5441	8,26713
% Akademiker	4,31452	2,45508
Mittleres Einkommen	102,835	18,7871
Varianz des Einkommens	2,89716	1,04771
% Ausländer	7,93337	4,09576
<b>Kreistypen</b>		
<i>Großer Verdichtungsraum</i>		
1: Kernstädte	<i>Referenzkat.</i>	<i>Referenzkat.</i>
2: hochverdichtet	0,14074	0,34780
3: verdichtet	0,092593	0,28990
4: ländlich	0,033333	0,17953
<i>Verdichtungsansatz</i>		
5: Kernstädte	0,077778	0,26786
6: verdichtet	0,22963	0,42065
7: ländlich	0,13704	0,34393
<i>ländlich geprägt</i>		
8: verdichtet	0,092593	0,28990
9: ländlich	0,051852	0,22176

In der folgenden Analyse werden Messwiederholungsmodelle geschätzt, bei dem die einzelnen Messzeitpunkte einer Region als analytische Ebene unterhalb der Ebene der Region betrachtet werden (vgl. *Engel* 1998: 116). Dabei können sowohl die Regressionskonstante als auch die Einflüsse der Zeit (in Jahren bzw. in quadrierten Jahren seit 1984) Zufallseffekte auf der Ebene der Regionen aufweisen. Zu beachten ist, dass der substanzielle Aussagegehalt dieser Modellierung auf der Aggregatenebene angesiedelt ist und Aussagen über individuelle Akteure immer mit der Gefahr des ökologischen Fehlschlusses behaftet sind. Formal ist das Modell für Ein- und Austritte in den Gleichungen 1 und 2 dargestellt. In den Gleichungen sind  $E_{ij}$ ,  $A_{ij}$  und  $N_{ij}$  die absoluten Zahlen der Eintritte, Austritte sowie der Beschäftigten zum



Zeitpunkt  $t$  in Region  $j$ . Die Zufallseffekte  $u_{0j}$ ,  $u_{1j}$  und  $u_{2j}$  beschreiben die regional spezifischen Abweichungen der Regressionskonstanten,<sup>4</sup> des Effektes der Zeit  $T$  sowie der quadrierten Zeit  $T^2$  von deren Mittelwerten. Mit Hilfe der beiden Effekte  $T$  und  $T^2$  können nicht-lineare Einflüsse der Zeit modelliert werden.  $X'$  stellt einen Zeilenvektor von zeitkonstanten und zeitveränderlichen Kovariaten dar und  $B$  einen Vektor der zugehörigen Koeffizienten. Wichtig ist, dass in Gleichung (1) (Eintritte) sowohl der Effekt der Zeit als auch der Effekt der quadrierten Zeit zwischen den Regionen variieren kann. In Gleichung 2 hingegen geht die quadrierte Zeit nur als fixer Effekt in das Modell ein, weil sich der Zufallseffekt von  $T^2$  als nicht signifikant erwiesen hat (vgl. unten). Im Prinzip ist es möglich, über Polynome höherer Ordnung auch nicht monotone Verläufe zu modellieren, um beispielsweise auch zyklische Prozesse abzubilden (*Singer* and *Willet* 2003: 214). Dafür hat sich das Zeitfenster in dieser Analyse jedoch als zu kurz erwiesen.

Eintritte:

$$\ln \left( \frac{\frac{E_{ij}}{N_{ij}}}{1 - \frac{E_{ij}}{N_{ij}}} \right) = b_{0j} + u_{0j} + (b_{1j} + u_{1j})T_j + (b_{2j} + u_{2j})T_j^2 + BX' + e_{ij} \quad (1)$$

Austritte:

$$\ln \left( \frac{\frac{A_{ij}}{N_{ij}}}{1 - \frac{A_{ij}}{N_{ij}}} \right) = b_{0j} + u_{0j} + (b_{1j} + u_{1j})T_j + b_{2j}T_j^2 + BX' + e_{ij} \quad (2)$$

Der Vorteil der Modellierung in den Gleichungen 1 und 2 besteht darin, dass durch sie die Schätzung der Effekte der unabhängigen Variablen auf die abhängige Variable nicht nur zu einem (aus forschungspragmatischen Gründen häufig willkürlich gesetzten) Zeitpunkt erfolgt, sondern als Längsschnittanalyse sich über einen längeren Zeitraum hinweg erstreckt. In diesem Messwiederholungsdesign stellen die meisten der unabhängigen Variablen zeitveränderliche Kovariaten dar.

Auf Grundlage der Daten des IAB Regionalfiles wurden für die folgenden empirischen Analysen Zeitreihen auf Regionalebene gebildet, in denen für jede Region in jedem Jahr von 1984 bis 1997 die Informationen der abhängigen und der erklärenden Variablen enthalten sind. Man erhält auf diese Weise eine aggregierte Datei mit einer Struktur der „Time Series Cross Section Data“ (TSCS) (*Beck* and *Katz* 1995,

4 Die Zufallseffekte  $u_{0j}$ ,  $u_{1j}$  und  $u_{2j}$  haben nichts mit  $u_i$  und  $u_j$  des *Lowry*-Modells zu tun.

2001b, vgl. Abbildung A1 im Anhang). In der Datenmatrix entspricht jede Zeile einer Messung einer Region im jeweiligen Kalenderjahr. Weil für jede Region vollständige Daten für den Zeitraum von 1984 bis 1997 vorliegen, geht jede Region mit 14 Zeilen in die Datenmatrix ein. Das bedeutet jedoch einen Verstoß gegen die Annahme der Unabhängigkeit der Beobachtungen voneinander, was zu einer Unterschätzung der Standardfehler und dadurch vermehrt zu einem fälschlichen Verwerfen der Nullhypothese führen kann (*Kreft* and *deLeeuw* 1998: 10). Denn natürlich sind wiederholte Messungen derselben Regionen *nicht* unabhängig voneinander. Weil die Schätzung der Standardfehler im Datensatz der Messwiederholungsanalyse also nicht auf  $N$  unabhängigen Informationen basiert, müssen die Standardfehler, je nach Ausmaß der durch Merkmale der Regionen bedingten Ähnlichkeit der Messungen innerhalb einer Region, nach oben korrigiert werden. Gelöst wird dieses Problem in dem hier vorgestellten Modell durch Schätzung einer zusätzlichen Fehlerkomponente: zum einen existiert ein Störterm  $u_{kj}$  aufgrund unbeobachteter Heterogenität und Messfehler für jede Region  $j$  und jede unabhängige Variable  $k$ , zum anderen ein Störterm  $e_{tj}$  aufgrund der unbeobachteten Heterogenität der zeitabhängigen Messungen zu  $t$  innerhalb jeder Region  $j$  (*Singer* and *Willett* 2003: 252). Dies impliziert eine Mischung von fixen und zufälligen Effekten (vgl. unten), weshalb derartige Schätzverfahren unter der Kategorie der „Mixed Models“ behandelt werden (*Littell* u.a. 1996: 492). Nimmt man an, dass beide Fehlerkomponenten rein zufällig sind, kann eine Normalverteilung der Fehler zugrunde gelegt werden; es wird also angenommen, dass die Fehler der einzelnen Messungen  $e_{tj}$  annähernd normalverteilt sind, einen Erwartungswert von Null und eine durch das Modell zu schätzende Varianz  $\sigma_e^2$  aufweisen:

$$e_{tj} \sim N(0, \sigma_e^2)$$

Wie in den Gleichungen (1) und (2) ersichtlich ist, weist der Fehlerterm aufgrund der hierarchischen Datenstruktur ein komplexeres Muster auf, indem auch  $u_{0j}$ ,  $u_{1j}$  und  $u_{2j}$  als Fehlerkomponenten auf der Ebene der Beobachtungseinheiten (das sind hier die Regionen) geschätzt werden. Für diese Fehlerkomponenten gilt wiederum die Annahme der Normalverteilung mit Erwartungswerten von Null:

$$E(u_{0j}) = E(u_{1j}) = E(u_{2j}) = 0$$

Daraus folgt, dass

$$E(\beta_{0j}) = \beta_0, \quad E(\beta_{1j}) = \beta_1, \quad E(\beta_{2j}) = \beta_2 .$$

Besonderheiten dieses Modells lassen sich anschaulicher darstellen durch einfaches Ausmultiplizieren und Umstellen von Gleichung (1). Daraus ergibt sich Gleichung

(3), dessen vorderer Teil die fixen Effekte und dessen hinterer Teil die zufälligen Effekte enthält:

$$\ln \left( \frac{\frac{E_{ij}}{N_{ij}}}{1 - \frac{E_{ij}}{N_{ij}}} \right) = \underbrace{b_{0j} + b_{1j}T_j + b_{2j}T_j^2 + BX' + u_{0j}}_{\text{fixer Teil}} + \underbrace{u_{1j}T_j + u_{2j}T_j^2 + e_{ij}}_{\text{zufälliger Teil}} \quad (3)$$

Im hinteren Teil der Gleichung (3) ist zu sehen, dass die Fehler der Beobachtungseinheiten nicht, wie es das klassische OLS Regressionsmodell erfordert, unkorreliert mit den unabhängigen Variablen  $T$  und  $T^2$  sind. Folglich impliziert diese Korrelation, dass auch die Residualvarianz der abhängigen Variablen von Werten unabhängiger Variablen abhängt und das Modell damit heteroskedastisch ist (*Engel* 1998: 74).<sup>5</sup>

Aufgrund der Zeitreihenstruktur ist überdies zu prüfen, ob eine Autokorrelation der Fehler  $e_{ij}$  mit den Fehlern der jeweils vorangegangenen Messung besteht. Dies geschieht durch eine Regression, bei der die Werte der Residuen  $e_{ij}$  durch die Residuen  $e_{ij-1}$  erklärt werden (*Maddala* 2002: 234):

$$e_{ij} = \rho e_{ij-1} + \varepsilon_t \quad (4)$$

Ein signifikanter Wert für  $\rho$  weist auf das Vorliegen eines autoregressiven Prozesses erster Ordnung hin. Verzichtet man auf dessen Kontrolle, werden die Standardfehler der Koeffizienten unterschätzt (ebd.: 240).<sup>6</sup>

Die Fehlerkovarianzmatrix  $r$  der Fehler  $e_{ij}$  lässt sich in blockdiagonaler Form darstellen (*Singer* and *Willett* 2003: 250):

$$r \sim N \begin{pmatrix} \Sigma_r & 0 & 0 & \dots & 0 \\ 0 & \Sigma_r & 0 & \dots & 0 \\ 0 & 0 & \Sigma_r & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & 0 & \Sigma_r \end{pmatrix}$$

5 Die Schätzung dieser Modelle erfolgt in SAS proc mixed nach dem Residual/Restricted Maximum Likelihood Verfahren (vgl. dazu *Singer* and *Willett* 2003: 87ff, *Littell* et al. 1996: 498).

6 Die Autokorrelation beträgt nach einem Jahr  $\rho$ , nach zwei Jahren  $\rho^2$  und nach  $n$  Jahren  $\rho^n$ .

Für  $r$  wird eine Normalverteilung mit einem Erwartungswert von 0 angenommen ( $E(r) = 0$ ), wie es für Residuen üblich ist. Die Matrix  $r$  enthält für jede Untersuchungseinheit  $j$ , in diesem Fall für jede Region, in der Diagonalen das Element  $\Sigma_r$  des Residuums  $r$ . Blockdiagonal ist die Matrix, weil die Elemente von  $r$  wiederum aus Matrizen gebildet sind. Diese Matrizen  $\Sigma_r$  haben folgende Gestalt, wenn man bei der Schätzung von einem autoregressiven Prozess erster Ordnung ausgeht:

$$\Sigma_r = \begin{pmatrix} \sigma^2 & \sigma^2\rho & \sigma^2\rho^2 & \sigma^2\rho^3 & \dots & \sigma^2\rho^{13} \\ \sigma^2\rho & \sigma^2 & \sigma^2\rho & \sigma^2\rho^2 & \dots & \sigma^2\rho^{12} \\ \sigma^2\rho^2 & \sigma^2\rho & \sigma^2 & \sigma^2\rho & \dots & \sigma^2\rho^{11} \\ \sigma^2\rho^3 & \sigma^2\rho^2 & \sigma^2\rho & \sigma^2 & \dots & \sigma^2\rho^{10} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma^2\rho^{13} & \sigma^2\rho^{12} & \sigma^2\rho^{10} & \sigma^2\rho^{11} & \dots & \sigma^2 \end{pmatrix}$$

Anhand von  $\Sigma_r$  ist zu sehen, dass die Fehlervarianzen der Nichtdiagonalen jeweils mit dem nach Gleichung (4) empirisch zu schätzendem Wert für  $\rho$  zusammenhängen und darum nicht, wie es das OLS Modell verlangen würde, innerhalb einer Region voneinander statistisch unabhängig sind. In Messwiederholungsmodellen sind autoregressive Prozesse der ersten Ordnung (AR1) häufig anzutreffen, es lassen sich aber auch Prozesse höherer Ordnungen schätzen, um beispielsweise auch zeitversetzt auftretende Autokorrelationen abzubilden. Man erhält bei 14 Messzeitpunkten für jede der 270 Regionen die oben gezeigte symmetrische autoregressive 14 x 14 Matrix  $\Sigma_r$ .

Dagegen steht in  $r$  die  $\mathbf{0}$  für eine symmetrische 14 x 14 Matrix, deren Elemente nur aus Nullen bestehen, was die Annahme der Unabhängigkeit der Regionen voneinander impliziert.<sup>7</sup>

Eine Verallgemeinerung von  $\Sigma_r$  stellt die heteroskedastisch-autoregressive Matrix  $\Sigma_{rh}$  dar. Während in der Hauptdiagonalen von  $\Sigma_r$  durch  $\sigma^2$  konstante Varianzen symbolisiert sind, was eine Homogenität der Varianzen über die einzelnen Messungen einer Region hinweg bedeutet, ist diese Homogenitätsannahme in  $\Sigma_{rh}$  gelockert, indem für jeden Messzeitpunkt eine *spezifische* Residualvarianz geschätzt wird.

<sup>7</sup> Weil es sich hier um Regionen handelt, die sich durch räumliche Nähe beeinflussen können, ist die durch die  $\mathbf{0}$  als nicht Diagonalelement von  $r$  eingeführte Annahme allerdings eher restriktiv.

$$\Sigma_{rh} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_2\rho^2 & \sigma_1\sigma_4\rho^3 & \cdots & \sigma_1\sigma_{14}\rho^{13} \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_2\sigma_3\rho & \sigma_2\sigma_4\rho^2 & \cdots & \sigma_2\sigma_{14}\rho^{12} \\ \sigma_3\sigma_1\rho^2 & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_3\sigma_4\rho & \cdots & \sigma_3\sigma_{14}\rho^{11} \\ \sigma_4\sigma_1\rho^3 & \sigma_4\sigma_2\rho^2 & \sigma_4\sigma_1\rho & \sigma_4^2 & \cdots & \sigma_4\sigma_{14}\rho^{10} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{14}\sigma_1\rho^{13} & \sigma_{14}\sigma_2\rho^{12} & \sigma_{14}\sigma_3\rho^{11} & \sigma_{14}\sigma_4\rho^{10} & \cdots & \sigma_{14}^2 \end{pmatrix}$$

In der Hauptdiagonalen von  $\Sigma_{rh}$  erhält man nun zu jedem Zeitpunkt einen spezifischen Wert für  $\sigma_k^2$ . Hier ist  $k$  ein Laufindex für die 14 Messzeitpunkte. Allerdings erhöht sich durch Lockerung der Annahme der Homoskedastizität auch die Zahl der zu schätzenden Kovarianzparameter von 8 auf 21 (Modell Gl. 1) bzw. 5 auf 18 (Modell Gl. 2). Somit muss das heteroskedastisch-autoregressive Modell nicht zwangsläufig auch das bessere sein, sondern die Modellanpassungen sind anhand des BIC Maßes von *Raftery* zu vergleichen. Je kleiner der Wert des BIC, desto besser ist das Modell den Daten angepasst, wobei das Maß die zusätzlichen Freiheitsgrade der weniger restriktiven Modelle berücksichtigt (*Singer and Willet* 2003: 121).<sup>8</sup>

## 5 Ergebnisse

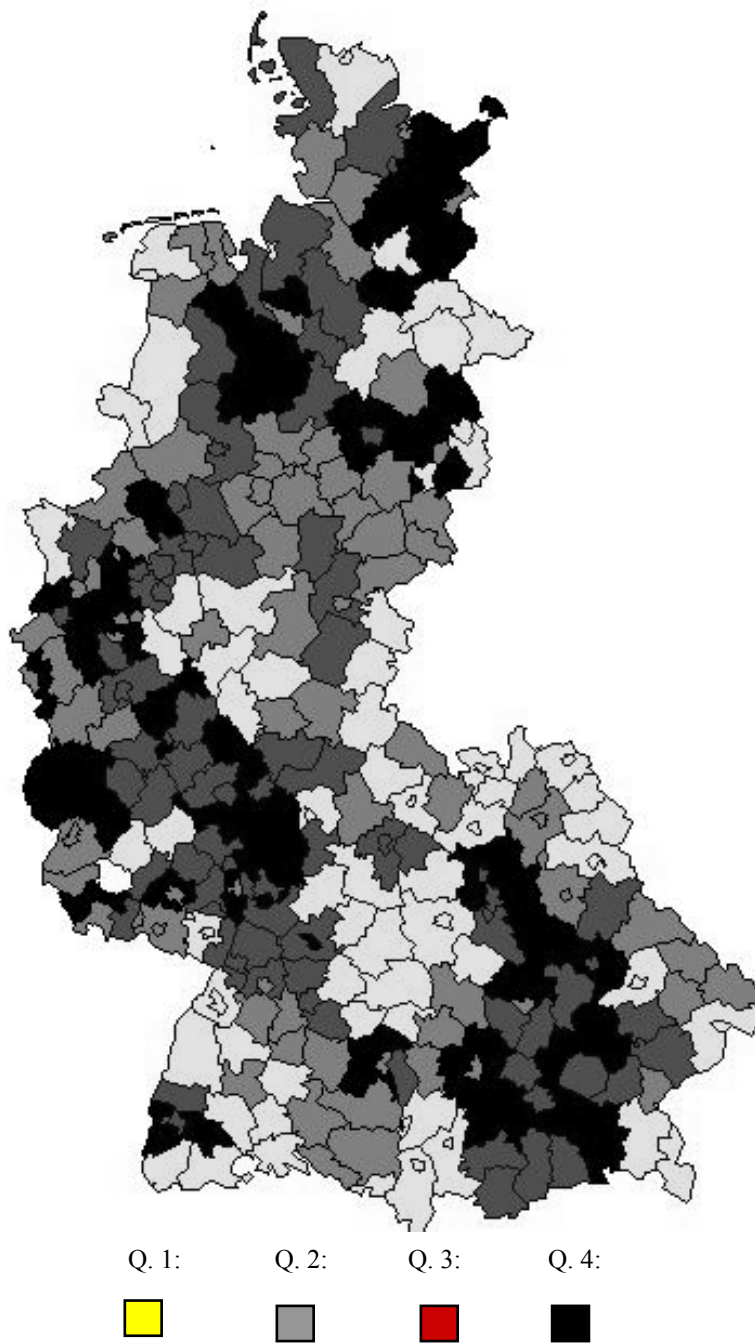
In der Abbildung 1 und der Abbildung 2 sind exemplarisch für das Jahr 1994 die regionalen Ein- und Austrittsquoten grafisch dargestellt.<sup>9</sup> Für jede Region wurden anhand der Größen  $E_{tj}$ ,  $A_{tj}$  und  $N_{tj}$  (vgl. Abschnitt 4) die jährlichen Ein- und Austrittsquoten  $E_{tj}/N_{tj}$  und  $A_{tj}/N_{tj}$  berechnet. Anschließend wurden anhand von Quartilen vier Kategorien von Regionen gebildet. In den Abbildungen deutet eine dunklere Färbung auf ein Quartil mit vergleichsweise hohen Ein- und Austrittsquoten einer Region hin. Hinsichtlich der Austrittsquoten lassen sich in Abbildung 1 grob drei Zentren feststellen: ein nördliches Zentrum um Hamburg, Bremen und Hannover, ein südliches Zentrum um München und Nürnberg, sowie eine etwas unzusammenhängende und weniger eindeutige Mitte-West-Region, die sich über das Neckar-

8 Noch flexibler wäre ein Modell mit einer völlig unstrukturierten Fehlerkovarianzmatrix, bei dem alle 105 Elemente der (symmetrischen) 14 x 14 Matrix  $\Sigma$  separat geschätzt werden. Hinzu käme in Modell 1 (Gl. 1) noch die Varianz-Kovarianzmatrix der Residuen der Ebene 2 der Regionen (für die Konstante und T, sowie T<sup>2</sup>) mit 6 Elementen, was zusammen 111 Varianz-Kovarianzparameter ergäbe. Die Flexibilität müsste man sich allerdings durch eine extrem hohe Zahl von Parametern erkaufen und es wäre wiederum durch das BIC Maß zu prüfen, ob man dadurch tatsächlich ein adäquateres Modell erhält. In dieser Analyse war jedoch für das Modell mit unstrukturierter Fehlerkovarianzmatrix keine stabile Lösung zu finden.

9 Aus dem Kreis Sankt Wendel (Saarland) gelangten keine Personen in die IAB-Regionalstichprobe, weshalb er aufgrund fehlender Werte als weiße Fläche dargestellt ist.

Rhein-Gebiet nach Nordwesten hin ausdehnend bis an den südlichen Rand des Ruhrgebietes erstreckt. Auch die meisten Kreise des Ruhrgebietes befinden sich im Quartil mit den höchsten Austrittsquoten. Sehr ähnlich verhält es sich bei den Eintrittsquoten in Abbildung 2. Herausragend ist sowohl bei den Eintritts- als auch bei den Austrittsquoten das südliche Zentrum um Nürnberg und München: Von 3780 Regionen-Jahren sind bei den Austritten 7 Kreise des Gebietes um München und 6 Kreise des Gebietes um Nürnberg (einschließlich Regensburg) unter den 20 höchsten Quoten. Bei den Eintrittsquoten sind von den 20 ersten Plätzen 15 von Kreisen um München besetzt, wobei 10 Plätze allein vom Landkreis München eingenommen werden.

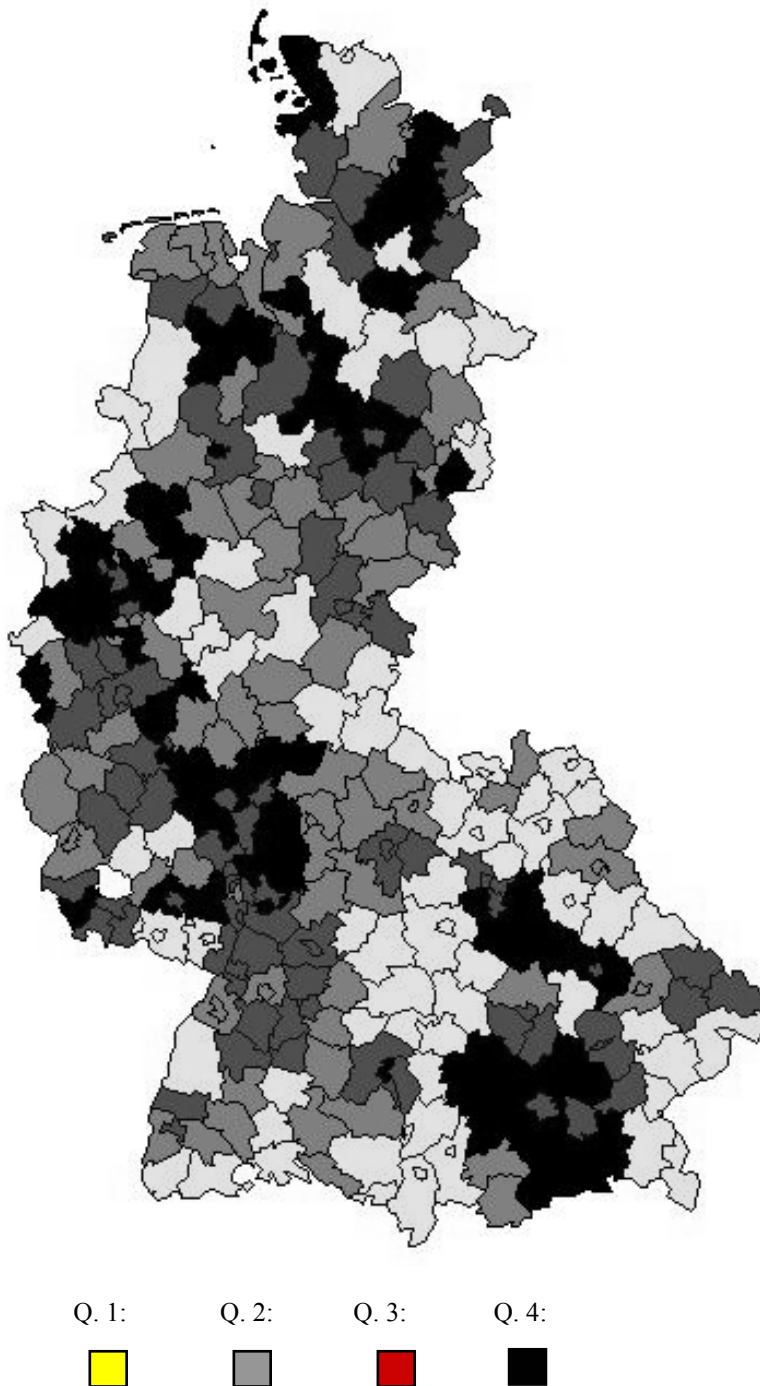
Die Tabelle 3 und die Abbildung 3 zeigen die Ergebnisse von Regressionsanalysen. In die in Tabelle 3 dargestellten OLS-Regressionsmodelle gingen die Zeiteffekte in Form von Dummytermen als erklärende Variablen ein. Die positiven Effekte der Zeit auf die Logits der Ein- und Austritte deuten auf einen leichten nicht-linearen Anstieg der Mobilität zwischen regionalen Arbeitsmärkten in den Jahren von 1984 bis 1997 hin. Über die logistische Linkfunktion  $\exp(\beta_0 + \beta_k) / [1 + \exp(\beta_0 + \beta_k)]$  wurden die Effekte auf die Logits in Anteilswerte zurückgerechnet und in Abbildung 3 grafisch dargestellt. Zu sehen ist dort, dass die Austrittsquote von ca. 6% 1984 auf knapp 8,5% im Jahr 1990 angestiegen ist und sich 1997 wieder dem Ursprungswert annäherte. Ähnlich verhält es sich mit der Eintrittsquote, obwohl diese im Jahr 1997 nicht in gleicher Weise wieder zurückgegangen ist.

**Abbildung 1** Austrittsquoten aus regionalen Arbeitsmärkten 1994, Quartile

Quelle: IAB Regionalstichprobe, eigene Berechnungen<sup>10</sup>

<sup>10</sup> Den Berechnungen der Quoten liegen die absoluten Zahlen der Beschäftigten (N) sowie die Ein- (E) und Austritte (A) einer Region in einem Jahr zu Grunde.

**Abbildung 2** Eintrittsquoten in regionale Arbeitsmärkte 1994, Quartile





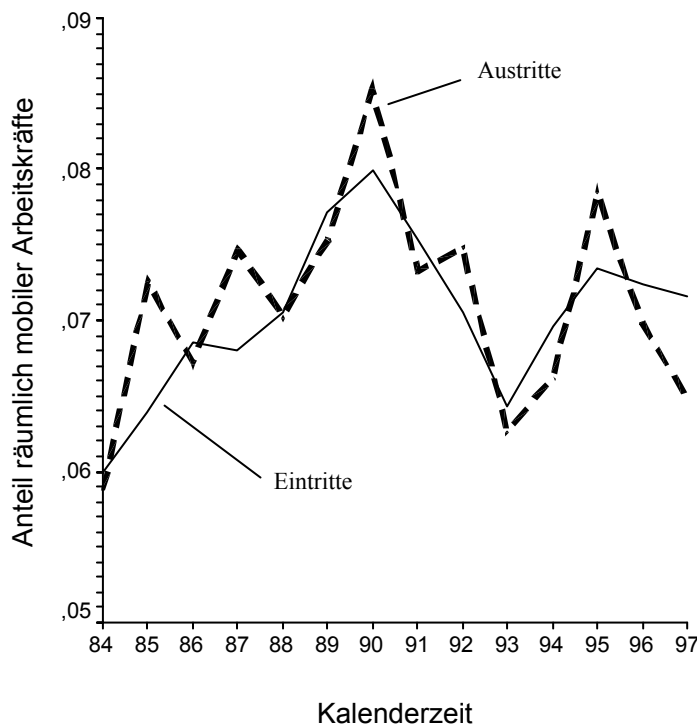
**Tabelle 3** Mobilität zwischen regionalen Arbeitsmärkten.  
Pooled OLS-Regression, Effekte der Kalenderzeit

Unabhängige Variablen: Zeiteffekte	AV: Logit(Eintritte)	AV: Logit(Austritte)
Konstante	-2.7538***	-2.7740***
Jahr 1984	<i>Referenzkat.</i>	<i>Referenzkat.</i>
Jahr 1985	0.0699*	0.2251***
Jahr 1986	0.1449***	0.1436***
Jahr 1987	0.1357***	0.2570***
Jahr 1988	0.1756***	0.1918***
Jahr 1989	0.2717***	0.2665***
Jahr 1990	0.3112***	0.4011***
Jahr 1991	0.2479***	0.2366***
Jahr 1992	0.1753***	0.2578***
Jahr 1993	0.0768*	0.0690*
Jahr 1994	0.1615***	0.1270***
Jahr 1995	0.2181***	0.3096***
Jahr 1996	0.2035***	0.1849***
Jahr 1997	0.1908***	0.1048***
N Regionen	270	
N Regionen-Jahre	3780	
R <sup>2</sup> (adjusted)	0.046 ***	0.085 ***

\* p<=0.05    \*\*p<=0.01    \*\*\*p<=0.001

Quelle: IAB Regionalstichprobe, eigene Berechnungen

**Abbildung 3** Verlauf der Ein- und Austrittsquoten über die Kalenderzeit



Der Verlauf der insgesamt leicht gestiegenen Quoten verdeutlicht, dass die Periodeneffekte nur ungenau durch eine metrische Zeitvariable eingefangen werden können, wenn diese einen linearen positiven Trend unterstellt. Insgesamt zeichnet sich zwar ein Anstieg der Mobilität in den untersuchten 14 Jahren ab, doch dieser Anstieg muss, wie noch zu sehen sein wird, differenziert betrachtet werden.

In Tabelle 5 wird anhand von Mehrebenenmodellen unter Kontrolle des autoregressiven Prozesses erster Ordnung gezeigt, durch welche regionalen Faktoren Eintritte und Austritte bestimmt sind. Die signifikanten Schätzungen von  $\rho$  deuten darauf hin, dass das Messwiederholungsdesign tatsächlich einen autoregressiven Prozess erster Ordnung beinhaltet. In Modell 2 (Austritte) wurde darüber hinaus eine Heterogenität der Residualvarianz über die Zeit hinweg zugelassen, nach dem die Auswahl der Modelle über einen Vergleich der BIC Werte getroffen wurde. Diese Werte sind in Tabelle 4 dargestellt.

**Tabelle 4** BIC Maße der Modelle

	homoskedastisch	heteroskedastisch
Abhängige Variable: Logit( <i>Eintritte</i> ) Modell 1	-1373,7	-1344,5
Abhängige Variable: Logit( <i>Austritte</i> ) Modell 2	-1203,5	-1212,3

Jeweils das Modell mit dem kleineren BIC Maß wurde ausgewählt, wobei der Unterschied bei den Austritten nach den Kriterien von *Raftery* (vgl. *Singer* and *Willet* 2003: 122) als „strong“, bei den Eintritten als „very strong“ angesehen werden kann (vgl. ebd.: 264, *Long* 1997: 112).

Modell 3 für Austritte in Tabelle 5 enthält alle drei Zufallseffekte – für die Konstante, für  $T$  und für  $T^2$ . Allerdings ist in Modell 3 nur der Zufallseffekt der Konstanten signifikant und darüber hinaus erwies sich die Matrix der Varianzen der Zufallseffekte als nicht positiv definit, weshalb die folgende Darstellung und Interpretation des Austrittsprozesses auf Modell 2 basiert.

Die signifikanten Varianzen der Zufallseffekte der Konstanten sowie des Effektes der metrischen Zeitvariablen  $T$  und  $T^2$  in den Modellen 1 und 2 in Tabelle 5 bedeuten, dass sich sowohl die mittlere Mobilitätsrate als auch der Einfluss der Zeit auf diese Rate signifikant zwischen den Regionen unterscheiden. Es bestätigt sich auch nach Kontrolle der zeitveränderlichen Einflussfaktoren sowohl für Ein- als auch für Austritte in der Grundtendenz der nichtlineare Zusammenhang aus Abbildung 3, da

der Zeiteffekt  $T$  positiv, der quadrierte Zeiteffekt  $T^2$  dagegen negativ ist. Aus dieser Kombination ergibt sich für die beobachtete Periode im Mittel ein im Zeitverlauf abnehmender Anstieg der Quoten. Die Modellierung mit zufälligen Koeffizienten hat den Vorteil, dass sich für jede Region ein eigenes Zeitprofil erstellen lässt. In den Abbildungen 4a-f sind die Profile der Eintrittsquoten für alle 270 Regionen grafisch dargestellt. Für diese Darstellung wurden auf Grundlage der Koeffizienten von  $T$  und  $T^2$  des Modells 1 in Tabelle 5 in Verbindung mit den Zufallseffekten die einzelnen Profile für den beobachteten Zeitraum von 14 Jahren über die logistische Linkfunktion  $\exp(Z)/[1+\exp(Z)]$  berechnet.<sup>11</sup> Als Intercept diente der in Tabelle 2 angegebene Mittelwert der Logits der Eintritte. Bei der Analyse der Austrittsquoten hat sich die Varianz des Zufallseffektes von  $T^2$  als nicht signifikant herausgestellt, so dass bei den Austritten die „Wölbung“ des Zeitprofils nicht zwischen den Regionen variiert. Darum sind die Profile der Austrittsquoten nicht grafisch dargestellt.

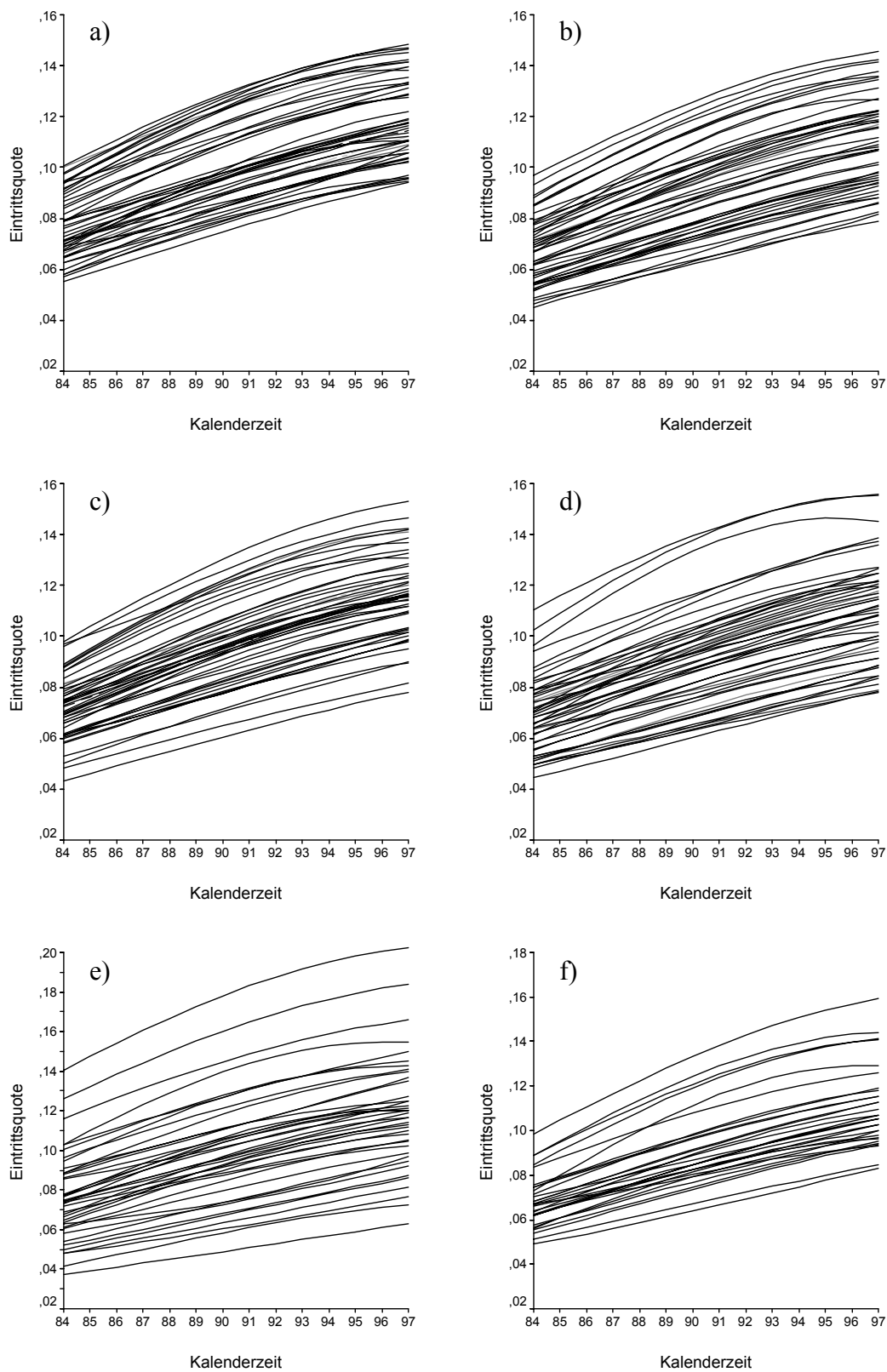
Einerseits zeigen die Abbildung 4a-f, dass in allen Regionen eine leicht ansteigende Tendenz der Eintrittsmobilität zu finden ist. Somit ist insgesamt festzuhalten, dass die in Abbildung 3 angedeutete Tendenz zum glockenförmigen Verlauf bei einem Großteil der Regionen auftritt. Zugleich zeigen sie andererseits aber auch, dass die Veränderungskurven in zweierlei Hinsicht variieren: erstens weist jede Kurve ein eigenes Ausgangsniveau der Eintrittsquote auf, wie anhand des jeweiligen Intercepts zu sehen ist. Darüber hinaus weist jede Kurve eine spezifische Steigung sowie eine spezifische Krümmung auf. Bei einigen ist die Verlaufsform exponentiell wachsend, was darauf hindeutet, dass auch die *Zunahme* der Mobilität in manchen Regionen im beobachteten Zeitraum gestiegen ist.

Immerhin zeigt die Abbildung 4a-f sehr deutlich, wie die einzelnen Profile voneinander abweichen und dass es sich daher lohnt, über die Berücksichtigung der Zufallseffekte die Zusammenhänge flexibel zu modellieren. Dies ist ein Vorteil des Random Coefficient Modells, auf den schon allein aus Gründen der korrekten Modellspezifikation nicht verzichtet werden sollte, auch wenn man von einer detaillierten „qualitativen“ Analyse der einzelnen regionalen Profile absieht.

Anhand der Abbildung 4a-f deutet sich bereits an, dass das Modell mit zufälligen Koeffizienten Beziehungen zu Analysen mit latenten Variablen aufweist. Denn die „hinter“ der Zeit und der quadrierten Zeit stehenden latenten Faktoren, die aufgrund der Zufallseffekte zu jeweils regional spezifischen Veränderungsprofilen führen, wurden nicht direkt gemessen.

---

11 Vgl. dazu *Rasbash* u.a. (2000: 146ff).

**Abbildung 4 a-f:** Zeitprofile der Eintrittsquoten in 270 Regionen

Quelle: IAB Regionalstichprobe, eigene Berechnungen, Schätzungen nach Modell I

Eine signifikante Variation der Zufallseffekte der Zeit weist also darauf hin, dass aufgrund nicht beobachteter Faktoren die Regionen unterschiedliche Veränderungsprofile aufweisen – ohne dass diese Faktoren jedoch inhaltlich konkretisiert werden.

Interessant ist nun der Effekt der Arbeitslosenquote. Für Eintritte ist dieser negativ und bestätigt damit die aus dem *Lowry*-Modell entwickelte Hypothese. Allerdings finden wir einen nicht signifikanten negativen Effekt der Arbeitslosenquote für Austritte, weshalb man nicht sagen kann, dass die überregionale Arbeitsmarktmobilität ein „push“ Effekt wäre. Der Abstrom aus den regionalen Arbeitsmärkten würde also umso *geringer* sein, je höher die regionale Arbeitslosenquote ist, wäre der Effekt signifikant.<sup>12</sup> Wollte man diesen der ökonomischen Theorie der räumlichen Mobilität tendenziell entgegengerichteten Befund tiefer gehend erklärend, wären Aussagen über die Mikroebene erforderlich, die auf Basis von Befunden einer Aggregatdatenanalyse grundsätzlich mit der Gefahr eines ökologischen Fehlschlusses behaftet sind. Zumindest lässt sich aber feststellen, dass die ökonomische Hypothese des *Lowry*-Modells einer empirischen Überprüfung nicht standhält (vgl. *Boyle, u.a.* 1998: 90).

In Hinblick auf den Test des ökonomischen *Lowry*-Modells ist festzustellen, dass die Quote der Austritte wie vorhergesagt mit zunehmenden mittleren Einkommen zurückgeht. Aus ökonomischer Sicht wird dies dadurch begründet, dass die nutzenmaximierenden Akteure Regionen mit hohem mittlerem Einkommen als attraktiv ansehen, wobei diese Begründung wiederum der Gefahr des ökologischen Fehlschlusses unterliegt. Allerdings ist ebenso für die Eintrittsquote ein negativer Effekt zu finden – was dem ökonomischen Argument entgegensteht.

Aus den Effekten der Dummyvariablen des Kreistyps lässt sich nach Kontrolle der Größe der Arbeitsmärkte herauslesen, dass sowohl die Eintritts- als auch die Austrittsraten in Regionen großer Verdichtungsräume am höchsten sind. Je peripherer die Lage ist, desto geringer ist tendenziell die überregionale Arbeitsmarktmobilität.

Insgesamt lässt sich also sagen, dass zum einen die überregionale Arbeitsmarktmobilität im beobachteten Zeitraum in der Tendenz leicht zugenommen hat, und dass zum anderen die ökonomischen Zusammenhänge, die dem *Lowry*-Modell zugrunde

---

12 Der Befund, dass die regionale Arbeitslosenquote keinen signifikanten positiven Effekt auf die Austrittsquote aufweist, erweist sich als sehr robust, wenn man in einer hierarchischen Modellbildung den Effekt der Arbeitslosenquote sukzessive um weitere Prädiktoren und Kontrollvariablen erweitert. Anders verhält es sich mit dem Effekt des mittleren regionalen Einkommens auf die Eintrittsquoten. Ist das mittlere regionale Einkommen der einzige Prädiktor im Modell, erhöht es die Eintrittsrate. Doch sobald die Zeit und die quadrierte Zeit kontrolliert sind, wird das Vorzeichen negativ.

**Tabelle 5** Mobilität zwischen regionalen Arbeitsmärkten. Random Coefficient Model Messwiederholungsanalyse von Aggregatdaten

Unabhängige Variablen: Merkmale der regionalen Arbeitsmärkte	Logit(Eintritte) Modell 1	Logit(Austritte) Modell 2	Logit(Austritte) Modell 3
<b>Fixe Effekte</b>			
Konstante	-1.2438 ***	-1.6029 ***	-1.6063 ***
T (Jahre seit 1984, metrisch) (t)	0.06005 ***	0.07301 ***	0.07278 ***
T <sup>2</sup> (Jahre seit 1984, metrisch, quadriert) (t)	-0.00156 ***	-0.00401 ***	-0.00402 ***
Ln(Beschäftigte in Region) (t)	-0.05926 *	-0.08744 **	-0.08814 **
% Beschäftigte wirtschafts- nahen Dienstleistungen (t)	0.007131 ***	0.006498 ***	0.006540 ***
Arbeitslosenquote (t)	-0.01777 ***	-0.00329 n.s.	-0.00331 n.s.
% Armut (HLU) (t)	-0.00173 ***	-0.00031 n.s.	-0.00034 n.s.
Distanz Umschlagbahnhof (min.)	0.000271 n.s.	0.000519 n.s.	0.000559 n.s.
Distanz Verdichtungsraum (min.)	-0.00445 ***	-0.00431 ***	-0.00434 ***
Δ Bevölkerung 1987-1997	0.01269 ***	0.009675 ***	0.009607 ***
% Akademiker (t)	0.02611 ***	0.01882 ***	0.01898 ***
Mittleres Einkommen (t)	-0.01017 ***	-0.00641 ***	-0.00633 ***
Varianz des Einkommens (t)	0.01589 *	0.01499 n.s.	0.01481 n.s.
% Ausländer (t)	-0.00865 **	-0.00467 n.s.	-0.00470 n.s.
<b>Kreistypen</b>			
<i>Großer Verdichtungsraum</i>			
1: Kernstädte	<i>Referenzkategorie</i>	<i>Referenzkategorie</i>	<i>Referenzkategorie</i>
2: hochverdichtet	-0.05609 n.s.	-0.01691 n.s.	-0.01537 n.s.
3: verdichtet	-0.1813 *	-0.1047 n.s.	-0.1036 n.s.
4: ländlich	-0.4064 ***	-0.2406 *	-0.2325 *
<i>Verdichtungsansatz</i>			
5: Kernstädte	-0.2347 ***	-0.1833 **	-0.1783 **
6: verdichtet	-0.4355 ***	-0.3004 ***	-0.2982 ***
7: ländlich	-0.4814 ***	-0.3445 ***	-0.3394 ***
<i>ländlich geprägt</i>			
8: verdichtet	-0.5221 ***	-0.3895 ***	-0.3820 ***
9: ländlich	-0.5126 ***	-0.3850 ***	-0.3819 ***
<b>Zufallseffekte</b>			
var(u <sub>0j</sub> )[Konstante]	0.06020 ***	0.05566 ***	0.05287 ***
var(u <sub>1j</sub> )[T]	0.000642 *	0.000086 ***	0.000092 n.s.
var(u <sub>2j</sub> )[T <sup>2</sup> ]	2.792E-6 *	--	2.836E-7 n.s.
cov(u <sub>0j</sub> ,u <sub>1j</sub> )	-0.00124 n.s.	-0.00116 ***	-0.00055 n.s.
cov(u <sub>0j</sub> ,u <sub>2j</sub> )	0.000021 n.s.	--	-0.00004 n.s.
cov(u <sub>1j</sub> ,u <sub>2j</sub> )	-0.00004 *	--	-2.07E-6 n.s.
ρ (AR 1)	0.1427 ***	--	--
ρ (ARH 1)	--	0.05308 **	0.05095 *
σ <sup>2</sup> <sub>etj</sub> a)	0.03078 ***	0.032 *** a)	0.032 *** a)
N Regionen = 270 (Level 2)			
N Regionen-Jahre = 3780 (Level 1)			
-2 Residual LL	-1418.09	-1313.11	-1314.05

\* p<=0.05      \*\*p<=0.01      \*\*\*p<=0.001

(t) = zeitveränderlich

a) bei den ARH 1-Modellen wurde der Mittelwert der heterogenen Level 1-Varianzen angegeben

Quelle: IAB Regionalstichprobe, eigene Berechnungen

liegen, allenfalls nur partiell auch für Westdeutschland gelten. Es deutet sich eher an, dass es regionale Faktoren gibt, welche die Zu- und Abstrommobilität auf den regionalen Arbeitsmärkten erhöhen oder reduzieren und deren Effekte sich nicht zwischen den abhängigen Variablen der Eintritts- und der Austrittsquoten unter-

scheiden. Zudem sind weitere Faktoren auf die überregionale Arbeitsmarktmobilität wirksam, von denen aus geographischer Sicht auch die infrastrukturellen Bedingungen zu zählen sind. Präzise theoretische Überlegungen zur Erklärung dieser geographischen Effekte müssen jedoch erst noch geleistet werden.

## 6 Diskussion des Modells mit zufälligen Koeffizienten

Über die Frage, inwieweit Modelle mit zufälligen Koeffizienten auch der explorativen Datenanalyse dienen können, wie in dieser Arbeit durch die Abbildungen 4 a-f angedeutet ist, existieren in der Literatur unterschiedliche Auffassungen (**Kreft** 1996: 8). Einige Autoren gehen davon aus, dass das Modell explorativen Zwecken dienen kann. Beispielsweise hebt **Goldstein** (1995: 2) die Anwendung der Mehrebenenanalyse für Rankingverfahren hervor, bei denen z.B. Schulen hinsichtlich ihrer relativen Effizienz beurteilt werden.<sup>13</sup> Auch in der Hochschulforschung ist ein Ranking unter Verwendung der Zufallseffekte  $u_{0j}$  üblich.<sup>14</sup> Häufig wird dabei anhand der fixen und zufälligen Effekte unter Kontrolle von Einflussfaktoren der Ebene 1 der mittlere Wert der Schülerleistung oder der Evaluation der Lehre berechnet (vgl. **Snijders** and **Bosker** 1999: 61f). Faktisch impliziert das Ranking zumeist eine inhaltliche Interpretation der Zufallseffekte, weil es auf nachvollziehbare Ursachen abzielt, die zu Unterschieden zwischen den Kontexteinheiten führen. Denn ohne die Vermutung konkreter Ursachen der Positionierung wäre eine individuelle Entscheidung für oder gegen eine bestimmte Bildungseinrichtung auf Grundlage des Rankings kaum rational.

Dagegen beschränken andere Autoren die Modelle mit zufälligen Koeffizienten auf verbesserte Schätzungen der Parameter, auf das Testen von Hypothesen und auf die Zerlegung der Varianz zwischen und innerhalb der Einheiten der Ebene 2 (**Bryk** and **Raudenbush** 1992: 5). Eine inhaltliche Interpretation der Zufallseffekte  $u_{kj}$  ist dabei nicht vorgesehen, sondern sie werden als *rein zufällige* Störgrößen und damit unabhängig von den erklärenden Variablen betrachtet. Im Lichte dieser offenen Debatte sollte eine inhaltliche Interpretation der Zufallseffekte auch im Messwiederholungsmodell folglich mit erhöhter Vorsicht geschehen.

---

13 "(...) there is often considerable interest in the relative ranking of individual schools, using the performances of their students after adjusting for intake achievements. This can be done straightforwardly using a multilevel modelling approach" (ebd.). Für eine Anwendung im Bereich der Industrieökonomik vgl. **Windzio** (2002).

14 explizit **Engel** und **Kreckeler** (2001: 136).

Bedacht werden sollte darüber hinaus, dass es sich bei den regionalen Einheiten der Ebene 2 nicht um eine Stichprobe handelt, sondern um eine Vollerhebung westdeutscher Regionen, weshalb man die Anwendung von Signifikanztests auf der regionalen Ebene in Frage stellen könnte. Werden andererseits – wie in dieser Arbeit – eher allgemeine Theorien getestet, kann dies mit der Superpopulationstheorie begründet werden, der zufolge die vorliegenden Daten ein Sample aus einer theoretischen Superpopulation darstellen, aus der für den Test der Theorien Anwendungsfälle gezogen werden (vgl. *Blien* 2002: 24, Anm. 2).

Bei Analysen mit räumlichem Bezug ist zudem generell die Möglichkeit räumlicher Korrelationen der Residuen zu bedenken. Allerdings wäre eine sowohl räumliche als auch zeitliche Korrelationen berücksichtigende Kovarianzmatrix überaus komplex und ist in der Standardsoftware bislang nicht verfügbar.

Ein kritischer Aspekt des Messwiederholungsmodells mit zufälligen Koeffizienten besteht in den Annahmen über die Verteilung der Zufallseffekte  $u_{kj}$  auf der Ebene 2 der Regionen. Wie dargelegt, wird bei der Schätzung der Zufallseffekte angenommen, dass diese einer Normalverteilung entstammen – was empirisch nicht unbedingt gegeben sein muss (*Swamy* 1970: 322). Aus diesem Grund geht ein neuerer nicht-parametrischer Ansatz, der auch bei Messwiederholungsmodellen angewandt wird, von der latenten Klassenanalyse aus (*Vermunt* and *Liesbet* 2001). Im Gegensatz zum klassischen parametrischen Ansatz des Modells mit zufälligen Koeffizienten, bei dem man für jede Einheit der Ebene 2 (hier: für jede Region) ein eigenes, einer multivariaten Normalverteilung entstammendes Residuum schätzt, wird beim nicht-parametrischen Ansatz eine finite Anzahl von Subgruppen angenommen, die als *latente Klassen* bezeichnet werden.<sup>15</sup> Nur für die latenten Klassen, nicht aber für jede einzelne Beobachtung der Ebene 2 erfolgt nun die Schätzung der Zufallseffekte, wodurch die Annahme normalverteilter  $u_{kj}$  der Ebene 2 nicht mehr benötigt wird. In der Forschungspraxis ist künftig wohl eine breitere Anwendung nicht-parametrischer Messwiederholungsmodelle zu erwarten.

## 7 Zusammenfassung und Schlussfolgerung

Die dargestellten Analysen haben gezeigt, dass es in *zeitlicher* Hinsicht zwar insgesamt zu einem leichten Anstieg der überregionalen Mobilität gekommen ist, dieser Anstieg selbst hat in den meisten Fällen jedoch abgenommen.

---

<sup>15</sup> In der Praxis wird die Anzahl der latenten Klassen sukzessive erhöht, bis sich die Modellanpassung nicht mehr verbessert (*Vermunt* and *Liesbet* 2001: 7).



Außerdem haben die deskriptiven Befunde ergeben, dass auch in räumlicher Hinsicht erwartungsgemäß eine hohe Varianz sowohl der Eintritts- als auch der Austrittsquoten existiert. Sowohl im Norden als auch in der Mitte sowie in Süddeutschland sind die höchsten Quoten der Ein- und Austritte auf Regionen konzentriert, die um urbane Zentren herum gelagert sind. Relativ eindeutig lässt sich sowohl bei den Ein- als auch bei den Austrittsquoten ein Gebiet als Spitzenreiter feststellen: nämlich die Regionen um Nürnberg und München, wobei wiederum dem Landkreis München selbst eine herausragende Stellung zukommt.

Darüber hinaus zeigte das Random Coefficient Messwiederholungsmodell, dass nicht alle Regionen dasselbe zeitliche Veränderungsprofil aufweisen und sich daher nicht alle Profile durch zwei fixe Koeffizienten  $T$  und  $T^2$  hätten abbilden lassen. Bei einer Grundtendenz von in abnehmendem Maße steigenden Quoten finden sich einige Regionen, die ein davon abweichendes Muster, etwa in Form von exponentiell steigenden Eintrittsquoten, aufweisen. Das Random Coefficient Modell mit fixen *und* zufälligen Effekten erwies sich als vorteilhaft, weil es hinreichend flexibel ist, um regionale Spezifika abzubilden.

Schließlich hat sich in der empirischen Analyse dieser Arbeit herausgestellt, dass in Regionen mit hoher Arbeitslosigkeit die Eintrittsquote von Beschäftigten gering ist, wie es nach dem *Lowry*-Modell zu erwarten war. Jedoch wurde andererseits deutlich, dass die Austrittsquote aus Regionen mit hoher Arbeitslosenquote *nicht* erhöht ist, was den Vorhersagen des *Lowry*-Modells widerspricht. Ähnlich verhält es sich mit dem mittleren regionalen Einkommen, welches sowohl die Ein- als auch die Austrittsquote reduziert, nach dem *Lowry*-Modell jedoch die Eintrittsquote hätte erhöhen müssen. Diese Befunde können durchaus als Einwand gegen das Modell angesehen werden, wengleich die getrennte Modellierung von Ein- und Austritten von *Lowrys* ursprünglicher Analyse der Mobilität *zwischen* Regionen abweicht.

In methodischer Hinsicht sollte deutlich geworden sein, dass das Random Coefficient Messwiederholungsmodell eine sehr flexible Variante der Modellierung darstellt. Was in dieser Analyse nur anhand der beiden Zeitvariablen herausgearbeitet wurde, dass nämlich über die Einbeziehung von zufälligen Effekten jede räumliche Einheit ihr eigenes Veränderungsprofil entwickeln kann, lässt sich ganz analog auch auf weitere zeitveränderliche erklärende Variablen übertragen. So könnte man auch untersuchen, ob der Zusammenhang zwischen der regionalen Arbeitslosenquote und der Ein- oder Austrittsquote zwischen den einzelnen Regionen variiert.

In dieser Arbeit gingen nur die beiden Zeitvariablen als Zufallseffekte in die Schätzung ein und es wurde darüber hinaus nicht untersucht, welche manifesten Variab-

len für die Variation der Zeiteffekte zwischen den Regionen verantwortlich sind. Diese Faktoren blieben darum gleichsam *latente* externe Größen. Es wurde somit aufgrund der signifikanten Schätzungen der Varianzen von  $u_{1j}$  und  $u_{2j}$  der Hinweis auf die Existenz latenter Variablen geliefert, die unabhängig von den im Modell berücksichtigten Faktoren die Zeitprofile der Regionen beeinflussen. Daher existieren Bezugspunkte des Random Coefficient Modells zu den Strukturgleichungsmodellen. In dem Lehrbuch von *Singer* und *Willet* (2003: 266ff.) sind die Beziehungen zwischen den beiden Verfahren näher ausgeführt.

Das Random Coefficient Modell ist auch für andere wissenschaftliche Disziplinen von Interesse. Häufig wurde die TSCS Analyse bislang in der international vergleichenden Politischen Ökonomie eingesetzt. Bisher scheinen die möglichen Vorteile einer Modellierung mit zufälligen Koeffizienten dort noch kaum ausgereizt zu sein, wengleich in jüngerer Zeit die methodische Diskussion darüber zugenommen hat (*Beck* and *Katz* 2001a, 2001b, *Western* 1998). Erwägt man eine vermehrte Anwendung des Random Coefficient TSCS Modells etwa in der Politischen Ökonomie, sollte allerdings bedacht werden, dass die hier vorgestellte Untersuchung auf 270 westdeutschen Regionen mit jeweils 14 Messzeitpunkten basierte. Die international vergleichende Forschung beruht dagegen häufig auf Daten der OECD, die wesentlich weniger Nationen umfassen. Wie gut dieses Modell auch bei kleineren Fallzahlen anwendbar ist, müsste erst durch weitere Forschung gezeigt werden. Die Analysen von *Western* (1998) und *Beck* und *Katz* (2001a) sind in dieser Hinsicht durchaus viel versprechend und es wäre sicherlich auch in der Politischen Ökonomie von Interesse, in multiplen Regressionsmodellen mit Hilfe der zufälligen Koeffizienten zeitliche Veränderungsprofile spezifisch für jede Nation zu ermitteln.

## 8 Anhang

**Tabelle A1:** Untergliederung der Regionstypen in Kreistypen

Regionstyp	Kreistyp	Charakter des Kreistyps
Großer Verdichtungsraum	1	Kernstädte in Regionen mit großen Verdichtungsräumen
	2	Hochverdichtete Kreise in Regionen mit großen Verdichtungsräumen
	3	Verdichtete Kreise in Regionen mit großen Verdichtungsräumen
	4	Ländliche Kreise in Regionen mit großen Verdichtungsräumen
Verdichtungsansatz	5	Kernstädte in Regionen mit Verdichtungsansätzen
	6	Verdichtete Kreise in Regionen mit Verdichtungsansätzen
	7	Ländliche Kreise in Regionen mit Verdichtungsansätzen
Ländlich geprägt	8	Verdichtete Kreise in ländlich geprägten Regionen
	9	Ländliche Kreise in ländlich geprägten Regionen

Quelle: Haas (2001: 134)

**Abbildung A1:** Datenmatrix der Messwiederholungsanalyse

Region	wave	Y	$X_t$	$X$
1002	1	0.01	0.03	0.54
1002	2	0.05	0.05	0.54
1002	3	0.09	0.17	0.54
⋮	⋮	⋮	⋮	⋮
1002	14	0.7	0.19	0.54
1003	1	0.05	0.09	0.23
1003	2	0.07	0.07	0.23
1003	3	0.10	0.12	0.23
⋮	⋮	⋮	⋮	⋮
1003	14	0.09	0.09	0.23
1051	1	0.03	0.13	0.09
1051	2	0.06	0.11	0.09
1051	3	0.09	0.09	0.09
⋮	⋮	⋮	⋮	⋮
1051	14	0.4	0.21	0.09
⋮	⋮	⋮	⋮	⋮

**Abbildung A2:** SAS `proc mixed` Programmsyntax für Modelle 1 und 2

```

/**** Modell 1 ****/
proc mixed data=zwei noclprint covtest; /**** ,zwei' ist der Datensatz ***/
class region wave; /**** Variablen ,region' und ,wave' nominalskaliert ****/
/**** fixed part des Modells ****/
model logit_ein=T t_sq log_n Ant_die_wg100 Alo Hlu T_bahn T_draum Bev_8097
Ant_akad100 Eink_mean Eink_var Ant_aus100 K2 K3 K4 K5 K6 K7 K8 K9
/s ddfm=bw notest ;
/**** random part des Modells ****/
random intercept t t_sq/solution type=un subject=region g;
/**** ar(1) => autoregressiv erster Ordnung****/
repeated wave/type=ar(1) subject=region r;
run;

/**** Modell 2 ****/
proc mixed data=zwei noclprint covtest;
class Region wave;
model logit_aus=T t_sq log_n Ant_die_wg100 Alo Hlu T_bahn T_draum Bev_8097
Ant_akad100 Eink_mean Eink_var Ant_aus100
K2 K3 K4 K5 K6 K7 K8 K9/s ddfm=bw notest ;
random intercept t /solution type=un subject=region g;
/**** arh(1) => autoregressiv-heteroskedastisch erster Ordnung****/
repeated wave/type=arh(1) subject=region r;
run;

```

Die Regionen-Perioden Datenmatrix (vgl. Abb. A1) enthält in der Variablen `Region` die ID Nummer der Region, `wave` zeigt die chronologische Nummer der Messung bzw. im Panelmodell den Wellenindikator an. Das `model` Statement leitet die Gleichung des fixen Teils ein. Durch das `s` werden die Schätzungen der fixen Effekte ausgegeben und `ddfm=bw` bedeutet, dass die Freiheitsgrade der fixen Effekte nach der „between-within“-Methode bestimmt werden. Demzufolge basiert der t-Test für jeden zeitkonstanten fixen Effekt auf der Ebene der Region auf  $270 - 12 = 258$  Freiheitsgraden, wobei 270 die Zahl der Regionen und 12 die Zahl der auf der Ebene der Regionen (d.h. zeitkonstant) geschätzten fixen Effekte ist (inklusive des Intercept). Dagegen basieren die t-Tests der zeitveränderlichen erklärenden Variablen der Ebene 1 auf  $3780 - 270 - 10 = 3500$  Freiheitsgraden, dabei ist 10 die Zahl der zeitveränderlichen erklärenden Variablen der Ebene 1 (vgl. Tabelle 5).

Hinter dem `random` Statement werden die Variablen aufgeführt, deren Effekte zwischen jedem `subject` (das sind hier die Regionen) variieren dürfen. Weil jedes Element der Varianz-Kovarianzmatrix  $\sigma$  der Ebene 2-Residuen  $u_{kj}$  aus den Daten geschätzt wird, ist diese Matrix unstrukturiert, was durch die Option `type=un` spezifiziert ist. Anders die Fehlerkovarianzmatrix der Ebene 1 Residuen  $e_{ij}$ , die SAS `proc mixed` als `r` Matrix bezeichnet und die hinter dem `repeated <Wellenindikator>` Statement angegeben wird. Mit `type=arh(1)` ist diese Matrix in Modell 2 autoregressiv (1. Ordnung) und heteroskedastisch, wodurch man eine Annahme über die Fehler in das Modell einführt, die gegenüber einer völlig unstrukturierten Matrix wesentlich sparsamer ist.

## Literatur

**Beck, Nathaniel, Jonathan Katz** (1995): What to do (and not to do) with time-series-cross-section data in comparative politics. *American Political Science Review* 89: 634-643.

**Beck, Nathaniel, Jonathan N. Katz** (2001a): Random coefficient models for time series cross section data: The 2001 version. Paper presented at the 2001 Annual meeting of the Society for Political Methodology.

**Beck, Nathaniel, Jonathan N. Katz** (2001b): Time series cross section data: what have we learned in the past few years? In: *Annual Review of Political Science* 4: 271-93

**Blien, Uwe** (2002): Arbeitslosigkeit und Entlohnung auf regionalen Arbeitsmärkten. Theoretische Analyse, ökonometrische Methode, empirische Evidenz und wirtschaftspolitische Schlussfolgerungen für die Bundesrepublik Deutschland. Heidelberg: Physica-Verlag.

**Boyle, Paul, Keith Halfacree, Vaughan Robinson** (1998): Exploring contemporary migration. Harlow/Essex: Longman.

**Bry, Antony S., Stephen W Raudenbush** (1992): Hierarchical linear models. Application and data analysis methods. London: Sage.

**Engel, Uwe** (1998): Einführung in die Mehrebenenanalyse. Grundlagen, Auswertungsverfahren und praktische Beispiele. Wiesbaden: Westdeutscher Verlag.

**Engel, Uwe, Gaby Kreckeler** (2001): Studienqualität. Über studentische Bewertungen und Rankings von Studienfächern einer Universität. S. 121-176 in: Uwe Engel (Hrsg.): Hochschulranking. Zur Qualitätsbewertung von Studium und Lehre. Frankfurt/Main: Campus.

**Goldstein, Harvey** (1995): Multilevel Statistical Models. London: Edward Arnold.

**Greene, William H.** (2000): *Econometric Analysis*. New Jersey: Prentice Hall.

- Haas, Anette* (2001): Die IAB-Regionalstichprobe 1975-1997. *ZA-Information* 48: 128-141.
- Hox, Joop* (2002): *Multilevel Analysis. Techniques and Applications*. Mahwah, New Jersey: Erlbaum.
- Kittel, Bernhard, Hannes Winner* (2002): How reliable is pooled analysis in political economy? The globalization-welfare state nexus revisited. MPIfG Discussion Paper 02 / 3.
- Kreft, Ita* (1996): Are multilevel techniques necessary? An overview, including simulation studies. Unpublished paper, Los Angeles, California State University.
- Kreft, Ita, Jan De Leeuw* (1998): *Introducing Multilevel Modeling*. London: Sage.
- Littell, Ramon C., George A. Miliken, Walter W. Stroup, Russel D. Wolfinger* (1996): *SAS System for Mixed Models*. Cary: SAS.
- Long, Scott* (1997): *Regression models for categorical and limited dependent variables*. London: Sage.
- Maddala, G.S.* (2002): *Introduction to econometrics*. Third edition. Chichester: Wiley.
- Milne, William* (1991): The human capital model and its econometric estimation. S. 137-151 in: Stillwell/Congdon.
- Rasbash, Jon, William Browne et al.* (2000): *A user's guide to MLwiN*. Multilevel Project, Institute of Education, London.
- Singer, Judith D.* (1998): Using SAS `proc mixed` to fit multilevel models, hierarchical models, and individual growth models. In: *Journal of Educational and Behavioral Statistics* 24: 323-355.
- Singer, Judith D., John B. Willet* (2003): *Applied longitudinal data analysis. Modeling change and event occurrence*. Oxford: University Press.
- Snijders, Tom, Roel Bosker* (1999): *Multilevel Analysis. An introduction to basic and advanced modeling*. London: Sage.
- Stillwell, John, Peter Congdon* (1991): *Migration Models. Macro and Micro Perspectives*. London: Belhaven
- Struck, Olaf, Christoph Köhler* (2004): *Beschäftigungsstabilität im Wandel? Empirische Befunde und theoretische Erklärungen für West- und Ostdeutschland*. München: Hampp.
- Swamy, P. (1970)*: Efficient inference in a random coefficient regression model. *Econometrica* 38: 311-323.
- Szell, G. (Hrsg.)*: *Regionale Mobilität*, S. 229-241. Nymphenburg: Nymphenburger Verlagsbuchhandlung.
- Vermunt, Jeroen K., Liesbet A. van Dijk* (2001): A non-parametric random coefficient approach: the latent class regression model. In: *Multilevel Modelling Newsletter* 13: 6-13.
- Western, Bruce* (1998): Causal heterogeneity in comparative research: A bayesian hierarchical modelling approach. In: *American Journal of Political Science* 42: 1233-1259.
- Windzio, Michael* (2002): Die "ökologische Dynamik" der Sozialstruktur in Organisationen. Ein zweistufiges Mehrebenenmodell zum Zusammenhang von Organisationsauflösungen und Beschäftigtenmobilität, in: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 54: 506-533.