

Rendtel, Ulrich; Schwarze, Johannes

Working Paper — Digitized Version

Zum Zusammenhang zwischen Lohnhöhe und Arbeitslosigkeit: Neue Befunde auf Basis semi-parametrischer Schätzungen und eines verallgemeinerten Varianz-Komponenten-Modells

DIW Discussion Papers, No. 118

Provided in Cooperation with:

German Institute for Economic Research (DIW Berlin)

Suggested Citation: Rendtel, Ulrich; Schwarze, Johannes (1995) : Zum Zusammenhang zwischen Lohnhöhe und Arbeitslosigkeit: Neue Befunde auf Basis semi-parametrischer Schätzungen und eines verallgemeinerten Varianz-Komponenten-Modells, DIW Discussion Papers, No. 118, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:

<https://hdl.handle.net/10419/95845>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



Diskussionspapiere
Discussion Papers

Diskussionspapier Nr. 118

**Zum Zusammenhang zwischen Lohnhöhe
und Arbeitslosigkeit:
Neue Befunde auf Basis semi-parametrischer
Schätzungen und eines verallgemeinerten
Varianz-Komponenten-Modells**

von

Ulrich Rendtel und Johannes Schwarze

Die in diesem Papier vertretenen Auffassungen liegen ausschließlich in der Verantwortung des Verfassers und nicht in der des Instituts.

Opinions expressed in this paper are those of the author and do not necessarily reflect views of the Institute.

Deutsches Institut für Wirtschaftsforschung

Diskussionspapier Nr. 118

**Zum Zusammenhang zwischen Lohnhöhe
und Arbeitslosigkeit:
Neue Befunde auf Basis semi-parametrischer
Schätzungen und eines verallgemeinerten
Varianz-Komponenten-Modells**

von

Ulrich Rendtel und Johannes Schwarze

Berlin, Juli 1995

Deutsches Institut für Wirtschaftsforschung, Berlin
Königin-Luise-Str. 5, D-14191 Berlin
Telefon: 49-30 - 89 7 89-0
Telefax: 49-30 - 89 7 89-200

Zum Zusammenhang zwischen Lohnhöhe und Arbeitslosigkeit: Neue Befunde auf Basis semi-parametrischer Schätzungen und eines verallgemeinerten Varianz-Komponenten-Modells

Ulrich Rendtel und Johannes Schwarze
Deutsches Institut für Wirtschaftsforschung
Berlin

Juni 1995

1 Problemstellung

Blanchflower und Oswald (1990, 1992) entwickeln auf Basis der Effizienzlohntheorie ein Modell, das einen negativen nicht-linearen Zusammenhang zwischen der regionalen Höhe der Arbeitslosigkeit und dem Lohnniveau postuliert. Dieses Modell wurde inzwischen für verschiedene Länder empirisch getestet und die theoretischen Aussagen wurden dabei weitgehend bestätigt¹.

Für die Bundesrepublik wurden empirische Analysen zur Lohnkurvendiskussion von Gerlach und Wagner (1994) und Wagner (1994a, b) vorgelegt. Sie schätzen auf Basis individueller gepoolter Querschnittsdaten aus mehreren Beobachtungsjahren eine aus der Humankapitaltheorie abgeleitete Lohngleichung, die sie um regionale Arbeitslosenquoten ergänzen. Für den Zeitraum 1979/1985 können Gerlach und Wagner (1994) mit Daten des BIBB einen signifikanten nicht-linearen Zusammenhang zwischen Arbeitslosigkeit und Lohnhöhe nachweisen. Dieser Nachweis gelingt jedoch nicht mit Daten des Sozio-oekonomischen Panels (SOEP) für den Zeitraum von 1984 bis 1990. Gerlach und Wagner (1994)

*Unser ganz besonderer Dank gilt Ulrich Pötter, der uns bei den Anwendungen semi-parametrischer Methoden eine große Hilfe war. Zu danken haben wir auch Bernd Adam, der das LaTeX Manuskript erstellt hat.

¹Vgl. für einen Überblick Blanchflower und Oswald (1990, 1992, 1994) oder Groot et al. (1992).

konnten Arbeitslosenraten allerdings nur auf der Ebene von Bundesländern verwenden, die keine theoretisch sinnvolle Abgrenzung regionaler Arbeitsmärkte darstellen. Wagner (1994b) schätzt Lohnkurven auf Basis von Beschäftigtendaten aus niedersächsischen Betrieben, die er mit Informationen zur Arbeitslosigkeit der niedersächsischen Raumordnungsregionen verbindet. Jedoch kann auch mit diesen Daten kein signifikanter negativer Zusammenhang zwischen Arbeitslosigkeit und Lohnhöhe festgestellt werden.

Bevor damit die Lohnkurvendiskussion für Deutschland vorschnell ad acta gelegt wird, möchten wir einige ergänzende und zum Teil weiterführende Analysen auf Basis des SOEP zu dieser Diskussion beisteuern. Da es inzwischen möglich ist, das SOEP auf kleinräumiger Ebene zu regionalisieren, verwenden wir Arbeitslosenquoten auf der Basis der 75 Raumordnungsregionen Westdeutschlands, durch die regionale Arbeitsmärkte sicherlich besser abgebildet werden, als durch Bundesländer.

Aus dem theoretischen Lohnkurven-Modell läßt sich nicht eindeutig ableiten, wie der nicht-lineare Zusammenhang zwischen Lohnhöhe und Arbeitslosigkeit funktionalisiert werden soll. Daher werden bei Schätzungen mit Individualdaten im allgemeinen mehrere ad-hoc Spezifikationen des Zusammenhangs geprüft. Eine Möglichkeit, dem "wahren" Zusammenhang auf die Spur zu kommen, ist die Verwendung semi-parametrischer Verfahren, die der Funktionsform einen größeren Spielraum lassen. Dazu möchten wir einige Ergebnisse vorstellen.

Ein anderes Problem der empirischen Lohnkurven-Analyse ist die Kontrolle von nicht beobachteten regionalen Effekten, die notwendig ist, um die theoretischen Kernaussagen des Modells testen zu können. Die übliche Vorgehensweise besteht darin, Querschnittsdaten aus mehreren Jahren zu poolen und fixe unbeobachtete Regionseffekte dann durch Dummy-Variablen für jede Region zu kontrollieren.

In einer ersten Erweiterung dieses Standardansatzes schätzen wir mit individuellen Paneldaten ein Error-Component-Modell, um damit die bekannten Probleme des Pooled-Regression-Ansatzes zu umgehen. Jedoch tauchen dann schnell Schätzprobleme auf, wenn unbeobachtete Regionseffekte für eine große Anzahl von Regionen durch Dummy-Variable kontrolliert werden sollen. Paneldaten bieten dazu jedoch eine andere Möglichkeit. Wir möchten eine Modellvariante benutzen, in der unbeobachtete regionsspezifische Effekte durch eine zusätzliche Varianzkomponente identifiziert werden. Voraussetzung dafür ist allerdings, daß man eine genügend große Zahl von Regionswechslern beobachten kann.

2 Theoretische Grundlagen der Lohnkurve

Einem negativen Zusammenhang zwischen Lohnhöhe und Arbeitslosigkeit sind in Deutschland durch flächendeckende Tarifverträge Grenzen gesetzt, die jedoch, wie das Beispiel Ostdeutschland zeigt, immer durchlässiger zu werden scheinen.

Aber selbst wenn Tarifverträge eingehalten werden, bleibt ein Spielraum für eine betriebsspezifische Lohndifferenzierung in Form der Lohndrift nach oben, also der Differenz zwischen Tarif- und Effektivlöhnen. Zahlreiche empirische Untersuchungen zu regionalen Effektivlohndifferenzen zeigen, daß dieser Spielraum auch genutzt wird (z.B. Jacoby 1990, Wagner 1991 oder Gornig und Schwarze 1992). Die eigentlichen Ursachen dafür konnten jedoch bisher nicht schlüssig geklärt werden.

Blanchflower und Oswald (1990, 1992) entwickeln ein Effizienzlohnmodell, das einen Zusammenhang zwischen der regionalen Höhe der Arbeitslosigkeit und dem Lohnniveau postuliert und geraten damit in Widerspruch zu anderen mikroökonomischen Ansätzen. Hall (1972) und Topel (1986) leiten auf der Basis der Theorie kompensierender Lohndifferentiale einen positiven Zusammenhang zwischen der regionalen Lohnhöhe und der Arbeitslosigkeit ab. Die Beschäftigten wählen diejenige Region aus, die ihren erwarteten Nutzen maximiert. Entscheidungsparameter sind der Lohn und die regionsspezifische Wahrscheinlichkeit, arbeitslos zu werden. Je höher diese Wahrscheinlichkeit ist, desto höher muß der gezahlte Lohn sein. Einen positiven Zusammenhang zwischen Lohnhöhe und Arbeitslosigkeit legt auch die Migrationstheorie nahe (vgl. Harris und Todaro 1970).

Im Effizienzlohnmodell von Blanchflower und Oswald produzieren die Unternehmen auf unvollkommenen Märkten und agieren dort als Preissetzer². Eine Konkurrenz zwischen Beschäftigten und Arbeitslosen ist nur eingeschränkt möglich, da die Arbeitslosen nicht in der Lage sind, sich durch Lohnunterbietung Arbeitsplätze zu beschaffen. Beide Annahmen dürften nicht unrealistisch sein und insbesondere die zweite Annahme wird in der Bundesrepublik durch flächendeckende Tarifverträge gestützt, die eine Lohnunterbietung durch Arbeitsanbieter weitgehend ausschließt.

Aufgrund ihrer Marktmacht erzielen die Unternehmen eine Rente, an der die Beschäftigten durch die Etablierung von Effizienzlohnmodellen beteiligt werden. Die Verhandlungsposition der Unternehmen ist durch die Möglichkeit der Entlassung gekennzeichnet, während die Arbeitnehmer die Möglichkeit des Streiks nutzen oder die Intensität ihrer Arbeitsleistung durch "Shirking" variieren können³.

Blanchflower und Oswald (1992) entwickeln auf Basis dieser Annahmen ein

²Zunächst hatten Blanchflower und Oswald (1990) das Modell als ein allgemeines betriebliches Lohnverhandlungsmodell entworfen, sind davon jedoch aus verschiedenen Gründen später abgerückt (vgl. dazu Blanchflower und Oswald 1992). Auch für Deutschland muß ein allgemeines Bargainingmodell, das auch die Verteilung der aus der Marktmacht der Unternehmen erwachsenden Rente einschließt, als nicht sehr plausibel abgelehnt werden. Die Tarifverhandlungen und deren Ergebnisse sind in Deutschland flächendeckend, während die Verteilung der unternehmensspezifischen Renten über die Lohndrift in den Betrieben entschieden wird. Die Annahme eines Effizienzlohnmodells scheint also plausibel.

³In Deutschland sind Streiks im Hinblick auf die Beteiligung an betriebsspezifischen Renten allerdings weitgehend ausgeschlossen. Sie sind ausdrücklich als "Verhandlungsmittel" des tariflichen Lohnbestandteiles und nicht als Mittel zur "Verteilung" unternehmensspezifischer Renten gedacht.

Modell für eine Volkswirtschaft mit zwei Regionen, das aber auf eine endliche Anzahl von Regionen ausgedehnt werden kann. In Region 1 sei der aus der Beschäftigung erzielte Nutzen u durch das Einkommen w_1 und die Anstrengung bei der Arbeit $e \in [0, 1]$ bestimmt. Der Nutzen jedes Beschäftigten läßt sich als Differenz von w_1 und e darstellen: $u_1 = w_1 - e$. Diese Nutzenfunktion impliziert, daß die Beschäftigten risikoneutral sind⁴. Der Arbeitseinsatz kann durch "Shirking", also durch verringerte Arbeitsintensität, reduziert und so der Nutzen erhöht werden, für den im Grenzfall $u = w_1$ gilt.

Shirking kann allerdings durch das Unternehmen entdeckt und der Beschäftigte daraufhin entlassen werden. Die Wahrscheinlichkeit entdeckt zu werden sei δ . Der erwartete Nutzen eines entlassenen Beschäftigten ist dann:

$$Eu_1 = (w_1 - e)\alpha(U_1) + b[1 - \alpha(U_1)]$$

Die Wahrscheinlichkeit α , einen neuen Arbeitsplatz zu finden ist eine Funktion in Abhängigkeit von der lokalen Arbeitslosigkeit U . b ist das Einkommensniveau, das durch die Arbeitslosenversicherung garantiert wird. Der Nutzen im Falle von Arbeitslosigkeit ist somit $u_1 = b$.

Zentral für die Ableitung der späteren Ergebnisse ist ein monoton fallender Verlauf der Funktion $\alpha(U)$, d.h. die Chance einer Wiederbeschäftigung sinkt mit steigender Arbeitslosigkeit. Deshalb soll die von Blanchflower und Oswald (1992) für diesen Zusammenhang gelieferte Begründung hier skizziert werden. Blanchflower und Oswald argumentieren, daß sich im langfristigen Gleichgewicht die Entlassungen und Neueinstellungen in einer Region entsprechen müssen. Die Entlassungen ergeben sich aus einer konstanten Quote von Unternehmenszusammenbrüchen r , multipliziert mit der regionalen Beschäftigung n . Neueinstellungen in der Region seien durch eine Funktion $\alpha[m - n]$ gekennzeichnet, mit m als der erwerbsfähigen Bevölkerung. Im langfristigen Gleichgewicht gilt dann: $rn = \alpha[m - n]$. Die regionale Arbeitslosenquote läßt sich als $U = (m - n)/m$ approximativ darstellen. Nach Auflösung dieser Bestimmungsgleichung nach n und Einsetzen in die Gleichgewichtsbedingung, erhält man eine Funktion für α in Abhängigkeit von U : $\alpha(U) = (r/U) - r$, mit $\alpha'(U) = -r/U^2 < 0$ und $\alpha''(U) = 2r/U^3 > 0$. Die Wahrscheinlichkeit einer erfolgreichen Arbeitssuche ist damit eine monoton fallende konvexe Funktion in Abhängigkeit von der Arbeitslosenquote⁵.

⁴Wobei für eine erste Ableitung der Nutzenfunktion nach w , $u_w > 0$ und für die erste Ableitung nach e , $u_e < 0$ gilt. Risikoneutralität liegt vor, da die zweiten Ableitungen der Nutzenfunktion gleich Null sind.

⁵In keiner der Arbeiten von Blanchflower und Oswald wird dieser Schritt explizit vollzogen. Er müßte sich aber wie folgt darstellen: Zunächst löst man die Bestimmungsgleichung für Arbeitslosigkeit $U = (m - n)/m$ nach n auf: $n = m(1 - U)$. Dieses Ergebnis setzt man in die Gleichgewichtsbedingung $rn = \alpha(m - n)$ ein und erhält: $rm(1 - U) = \alpha[m - m(1 - U)]$. Bei der Auflösung dieser Gleichung nach α entfällt m und es ergibt sich: $\alpha = (r/U) - r$.

Modell für eine Volkswirtschaft mit zwei Regionen, das aber auf eine endliche Anzahl von Regionen ausgedehnt werden kann. In Region 1 sei der aus der Beschäftigung erzielte Nutzen u durch das Einkommen w_1 und die Anstrengung bei der Arbeit $e \in [0, 1]$ bestimmt. Der Nutzen jedes Beschäftigten läßt sich als Differenz von w_1 und e darstellen: $u_1 = w_1 - e$. Diese Nutzenfunktion impliziert, daß die Beschäftigten risikoneutral sind⁴. Der Arbeitseinsatz kann durch "Shirking", also durch verringerte Arbeitsintensität, reduziert und so der Nutzen erhöht werden, für den im Grenzfall $u = w_1$ gilt.

Shirking kann allerdings durch das Unternehmen entdeckt und der Beschäftigte daraufhin entlassen werden. Die Wahrscheinlichkeit entdeckt zu werden sei δ . Der erwartete Nutzen eines entlassenen Beschäftigten ist dann:

$$Eu_1 = (w_1 - e)\alpha(U_1) + b[1 - \alpha(U_1)]$$

Die Wahrscheinlichkeit α , einen neuen Arbeitsplatz zu finden ist eine Funktion in Abhängigkeit von der lokalen Arbeitslosigkeit U . b ist das Einkommensniveau, das durch die Arbeitslosenversicherung garantiert wird. Der Nutzen im Falle von Arbeitslosigkeit ist somit $u_1 = b$.

Zentral für die Ableitung der späteren Ergebnisse ist ein monoton fallender Verlauf der Funktion $\alpha(U)$, d.h. die Chance einer Wiederbeschäftigung sinkt mit steigender Arbeitslosigkeit. Deshalb soll die von Blanchflower und Oswald (1992) für diesen Zusammenhang gelieferte Begründung hier skizziert werden. Blanchflower und Oswald argumentieren, daß sich im langfristigen Gleichgewicht die Entlassungen und Neueinstellungen in einer Region entsprechen müssen. Die Entlassungen ergeben sich aus einer konstanten Quote von Unternehmenszusammenbrüchen r , multipliziert mit der regionalen Beschäftigung n . Neueinstellungen in der Region seien durch eine Funktion $\alpha[m - n]$ gekennzeichnet, mit m als der erwerbsfähigen Bevölkerung. Im langfristigen Gleichgewicht gilt dann: $rn = \alpha[m - n]$. Die regionale Arbeitslosenquote läßt sich als $U = (m - n)/m$ approximativ darstellen. Nach Auflösung dieser Bestimmungsgleichung nach n und Einsetzen in die Gleichgewichtsbedingung, erhält man eine Funktion für α in Abhängigkeit von U : $\alpha(U) = (r/U) - r$, mit $\alpha'(U) = -r/U^2 < 0$ und $\alpha''(U) = 2r/U^3 > 0$. Die Wahrscheinlichkeit einer erfolgreichen Arbeitssuche ist damit eine monoton fallende konvexe Funktion in Abhängigkeit von der Arbeitslosenquote⁵.

⁴Wobei für eine erste Ableitung der Nutzenfunktion nach w , $u_w > 0$ und für die erste Ableitung nach e , $u_e < 0$ gilt. Risikoneutralität liegt vor, da die zweiten Ableitungen der Nutzenfunktion gleich Null sind.

⁵In keiner der Arbeiten von Blanchflower und Oswald wird dieser Schritt explizit vollzogen. Er müßte sich aber wie folgt darstellen: Zunächst löst man die Bestimmungsgleichung für Arbeitslosigkeit $U = (m - n)/m$ nach n auf: $n = m(1 - U)$. Dieses Ergebnis setzt man in die Gleichgewichtsbedingung $rn = \alpha(m - n)$ ein und erhält: $rm(1 - U) = \alpha[m - m(1 - U)]$. Bei der Auflösung dieser Gleichung nach α entfällt m und es ergibt sich: $\alpha = (r/U) - r$.

Die zweite Region ist durch den Lohnsatz w_2 und Arbeitslosigkeit U_2 gekennzeichnet. b und e sind hier identisch⁶. Darüberhinaus ist Region 2 durch bestimmte nicht-monetäre Attraktivitäten ϕ gekennzeichnet (man denke etwa an Regionen mit guten Freizeitmöglichkeiten). Der Nutzen eines Beschäftigten ist dort also $u_2 = w_2 - e + \phi$, der eines Arbeitslosen $u_2 = b + \phi$.

Die Nachfrage nach Arbeit wird durch regionsspezifische Schocks beeinflusst. Schließlich wird die Möglichkeit der Wanderung zwischen beiden Regionen eingeräumt, allerdings nur zwischen den Perioden und nicht während einer laufenden Periode. Das Gleichgewicht dieses Effizienzlohnmodells ist dadurch gekennzeichnet, daß der erwartete Nutzen einer Situation mit Shirking, dem erwarteten Nutzen einer Situation ohne Shirking entspricht. Dies ergibt sich aus der optimalen Lohnsetzung der Unternehmen einerseits - sie müssen einen ausreichend hohen Lohn zahlen, wenn sie Shirking verhindern wollen - und dem optimalen Verhalten der Beschäftigten im Hinblick auf ihren Arbeitseinsatz andererseits.

Aus dieser Gleichgewichtsbedingung kann eine für beide Regionen gleiche konvexe Lohnkurve mit fallender Steigung abgeleitet werden⁷:

$$w = e + b + \frac{e\delta}{(1 - \delta)[1 - \alpha(U)]}$$

Die Konvexität von w ist Resultat der Konvexität von $\alpha(U)$. Die zusätzliche Attraktivität der zweiten Region, spielt in der Lohngleichung keine Rolle, da sich ϕ herauskürzt. Da aber ϕ sehr wohl in der Nutzenfunktion von Beschäftigten (aber auch nicht Beschäftigten) eine Rolle spielt, folgt, daß die weniger attraktive Region 1 durch einen höheren erwarteten Lohnsatz gekennzeichnet sein muß. Beide Regionen müssen den Beschäftigten dasselbe erwartete Nutzenniveau bieten, da ansonsten der Wegzug sämtlicher Beschäftigter aus der weniger attraktiven Region erfolgen würde. Die Gleichgewichtssituation des Modells erzeugt also eine Situation, in der Wanderungen nicht mehr stattfinden.

Die Regionen müssen sich dann offensichtlich durch eine unterschiedliche Verteilung der Nachfrageschocks auszeichnen, verbunden mit unterschiedlichen Lohn/Arbeitslosigkeits-Kombinationen. Im Falle identischer Nachfrageschocks wäre in beiden Regionen die gleiche Lohn/Arbeitslosigkeits-Kombination zu beobachten. Da aber Region 2 attraktiver ist, also nicht-monetäre Vorteile bietet, würde dies Wanderungen aus der ersten Region induzieren. In einem Gleichgewicht ohne Wanderungen muß also die Attraktivität der zweiten Region durch eine ungünstigere Position auf der Lohn/Arbeitslosigkeits-Kurve ausgeglichen

⁶Blanchflower und Oswald (1992) analysieren auch den Fall unterschiedlicher Einkommensniveaus bei Arbeitslosigkeit. Bei einer Analyse für Deutschland, wo die Einkommensersatzleistung bei Arbeitslosigkeit bundesweit einheitlich geregelt ist, kann davon aber abgesehen werden.

⁷Dies gilt allerdings nur, wenn das Niveau des Arbeitslosengeldes in beiden Regionen gleich ist.

werden. Dies soll durch unterschiedliche Nachfrageschocks erreicht werden: "Repeated random shocks produce different points on the curve. To ensure a zero-migration equilibrium, the intrinsically more attractive region 2 must be characterized more often by points in the south-east portion of the wage curve, which implies worse wage and unemployment combinations". (Blanchflower und Oswald 1992, 8). Attraktive Regionen wären danach eher im süd-östlichen, weniger attraktive Regionen im nord-westlichen Teil des Lohnkurvendiagramms zu finden.

3 Stand der empirischen Forschung

Ökonometrisch wird das Effizienzlohnmodell häufig als ein Eingleichungsmodell spezifiziert, das mit OLS geschätzt wird⁸:

$$w_j = f(U_j, j)$$

Das Lohnniveau in Region j ist eine Funktion der regionalen Arbeitslosenquote und bestimmter regionaler Charakteristika j . Als Basismodell für die Schätzung der Lohnkurve wird i.d.R. eine auf individuellen Einkommensdaten beruhende Einkommensfunktion, z.B. vom Mincer-Typ, verwendet, um die individuen-spezifischen Bestimmungsfaktoren des Lohns, wie Qualifikation, Berufserfahrung und Arbeitszeit zu kontrollieren.

Die genaue Spezifikation des monoton fallenden Zusammenhangs zwischen Lohnhöhe und Arbeitslosigkeit kann aus der Theorie nicht abgeleitet werden. Deshalb wird die individuelle Einkommensgleichung durch verschiedene ad-hoc Spezifikationen der regionalen Arbeitslosenquote ergänzt, die den theoretisch postulierten Zusammenhang zwischen Lohnhöhe und Arbeitslosigkeit approximieren sollen⁹. Blanchflower und Oswald (1992, 1994), aber auch Groot et al. (1992) testen unterschiedliche Spezifikationen und gelangen zu dem Schluß, daß sich die quadratische (U, U^2), die logarithmische ($\log U$) und die log-kubische ($\log U, (\log U)^3$) Approximation als sinnvoll erweisen.

⁸Eine vollständige Spezifikation des Modells enthält noch zwei weitere Gleichungen. Eine für die Bestimmung der regionalen Arbeitslosigkeit und eine Bedingung die sicherstellt, daß der erwartete Nutzen in allen Regionen gleich ist und die zero-migration Bedingung wiedergibt. Blanchflower und Oswald (1992, 13) zeigen aber, daß man sich mit einer einfachen OLS-Schätzung der Lohngleichung immer auf der "sicheren Seite" befindet, da in diesem Fall der Koeffizient der Arbeitslosenquote nach oben verzerrt geschätzt wird und es somit schwierig ist, einen signifikanten negativen Koeffizienten zu beobachten.

⁹Für die Schätzung der Lohnkurve werden also Informationen auf unterschiedlichem Aggregationsniveau verwendet. Einerseits individuelle Daten, andererseits auf Regionsebene aggregierte Arbeitslosenquoten. Dabei ergibt sich als ein ökonometrisches Problem, daß der Einfluß der Arbeitslosenquote auf die Einkommen überschätzt wird, da die Standardabweichung des Parameters systematisch unterschätzt wird. Auf dieses Problem soll hier aber nicht weiter eingegangen werden (vgl. dazu Moulton 1990).

Besondere Bedeutung kommt nach Blanchflower und Oswald der Kontrolle der fixen regionsspezifischen Effekte β_j zu, deren Berücksichtigung bei der Schätzung sie als unerlässlich einstufen, da diese u.a. regionsspezifische Nachfragebedingungen, die unterschiedliche Attraktivität der Regionen sowie Unterschiede im Preisniveau kontrollieren. Letzteres ist von Bedeutung, da das Effizienzlohnmodell für Reallöhne formuliert ist, die jedoch empirisch nur sehr schwer zu beobachten sind. Sind die genannten regionsspezifischen Eigenschaften über einen längeren Zeitraum konstant, dann kann die Lohnkurve mit über mehrere Jahre gepoolten individuellen Einkommensdaten geschätzt - und so zumindest die fixen regionsspezifische Effekte durch Dummy-Variablen kontrolliert werden.

Die auf diesem Verfahren basierenden empirischen Arbeiten von Blanchflower und Oswald (1990, 1992, 1994) oder Groot et al. (1992) führen zu einer Bestätigung des Lohnkurvenmodells. Die im internationalen Vergleich geschätzten Elastizitäten des Lohnes bezüglich der Arbeitslosigkeit sind sich überraschend ähnlich und liegen im Durchschnitt bei -0.1, was bedeutet, daß eine Verdoppelung der Arbeitslosigkeit, die Lohnsätze um 10% senkt.

Für die westdeutschen Bundesländer wurden Analysen zur Lohnkurve von Gerlach und Wagner (1994) und Wagner (1994a, b) vorgelegt. Sie verwenden gepoolte Querschnittsdaten abhängig beschäftigter Männer, wobei sie Beamte und Beschäftigte im öffentlichen Dienst ausschließen, da deren Entlohnung bundesweit einheitlich geregelt ist und keinen Spielraum für eine "innerbetriebliche" Lohnbildung läßt.

Für den Zeitraum 1979/1985 (Datenbasis ist eine Umfrage des Bundesinstituts für Berufsbildung und des IAB) können Gerlach und Wagner (1994) bestätigen, daß es eine monoton fallende, nicht lineare Lohnkurve auch in Deutschland gibt. Unter Berücksichtigung fixer Effekte ermitteln sie eine Elastizität des Lohnsatzes bezüglich der Arbeitslosigkeit von -0.13, ein Ergebnis, das denen von Blanchflower und Oswald (1992) für viele andere Länder ermittelten, sehr ähnlich ist.

Für den Zeitraum von 1984 bis 1990 schätzen Gerlach und Wagner (1994) die Lohnkurve mit Daten des Sozio-oekonomischen Panels, ohne jedoch den Längsschnittcharakter dieses Datensatzes zu nutzen. Dabei können sie die Ergebnisse ihrer Schätzung mit den IAB-Daten jedoch nicht bestätigen. Bei Berücksichtigung fixer Regionaleffekte ist keine der Spezifikationen auf dem üblichen Niveau signifikant. Die gefundenen Unterschiede führen Gerlach und Wagner (1994) auf die veränderte Situation auf dem Arbeitsmarkt zurück: "Während in Zeiten einer geringen Arbeitslosigkeit zu Anfang der achtziger Jahre Firmen in Regionen mit einer verglichen zum nationalen Niveau noch geringen Arbeitslosenquote Effizienzlöhne zahlten, brauchten angesichts des hohen Arbeitslosigkeitsniveaus Firmen dies in der zweiten Hälfte der achtziger Jahre auch in Bundesländern mit einer verglichen zum nationalen Niveau geringen Arbeitslosenquote nicht zu tun." (Gerlach und Wagner 1994, 11).

Wagner (1994b) schätzt Lohnkurven für 1979 bis 1990 auf Basis von Beschäftigten- und Lohndaten niedersächsischer Industriebetriebe. Diese Daten verknüpft er mit der

Arbeitslosenquote der niedersächsischen Raumordnungsregionen. Wagner (1994b) kann weder für den gesamten Zeitraum, noch für Teilzeiträume einen monoton fallenden nicht-linearen Zusammenhang zwischen Lohnhöhe und Arbeitslosigkeit nachweisen.

4 Der Datensatz

Die in diesem Beitrag präsentierten Ergebnisse beruhen auf Daten des SOEP 1985, 1989 und 1993 (vgl. zum SOEP Projektgruppe Panel 1995). Diese Auswahl ist durch die Verfügbarkeit von Arbeitslosenquoten für Raumordnungsregionen bestimmt. Regionale Indikatoren werden von der Bundesanstalt für Landeskunde und Raumordnung (BfLR) in unregelmäßigen Abständen für die 75 Raumordnungsregionen der alten Bundesländer zur Verfügung gestellt¹⁰. Raumordnungsregionen bilden das in Oberzentren gegliederte Standortssystem der Bundesrepublik sowie die dazugehörigen Einzugsbereiche ab (vgl. BfLR 1995) und stellen somit auch eine sinnvolle Abgrenzung regionaler Arbeitsmärkte dar. Für unsere Analyse konnten wir auf originäre Arbeitslosenquoten für die Jahre 1989 und 1993 zurückgreifen. Da regionale Angaben zur Entwicklung der Arbeitslosigkeit von 1985 bis 1989 vorliegen, war auch eine Berechnung der Arbeitslosenquote für 1985 möglich. Die Arbeitslosenquoten für diese Jahre wurden mit dem aus dem SOEP gewonnenen Individualdatensatz verknüpft.

Der Individualdatensatz enthält ausschließlich Männer, die in mindestens zwei der drei analysierten Jahre in der Privatwirtschaft als Arbeiter oder Angestellte vollzeitbeschäftigt waren. Beschäftigte im öffentlichen Dienst werden aus den erwähnten Gründen von der Analyse ausgeschlossen. Das analysierte Einkommen ist das logarithmierte monatliche Bruttoerwerbseinkommen. Die individuellen lohnbestimmenden Faktoren werden durch eine Einkommensgleichung vom Mincer-Typ kontrolliert, wobei im einzelnen die folgenden Merkmale verwendet werden: Die Ausbildungsdauer in Jahren, die Jahre der Berufserfahrung, die Dauer der Betriebszugehörigkeit, die logarithmierte wöchentliche Arbeitszeit, die berufliche Stellung, sowie sechs Dummy-Variable die die Zugehörigkeit zu bestimmten Wirtschaftszweigen erfassen und zwei Dummy-Variable für 1989 und 1993, um zeitliche Einflüsse zu kontrollieren. Eine vollständige Übersicht über alle verwendeten Merkmale zeigt Tabelle 1.

¹⁰Seit kurzem sind solche Daten auch für die 22 Raumordnungsregionen der neuen Bundesländer verfügbar.

5 Ergebnisse der semi-parametrischen Modellspezifikation

Aus dem theoretischen Modell von Blanchflower und Oswald geht nicht hervor, wie die monoton fallende Lohnkurve ökonometrisch spezifiziert werden soll. Dieses Problem wurde bereits diskutiert. Bevor wir Regressionsmodelle mit einer festgelegten Spezifikation des Zusammenhangs schätzen, soll deshalb mit einer semi-parametrischen Methode untersucht werden, wie sich der Zusammenhang zwischen Lohnhöhe und Arbeitslosigkeit in den von uns verwendeten Daten darstellt.

Wir benutzen dazu verallgemeinerte additive Modelle, wie sie etwa bei Hastie und Tibshirani (1990) oder Green und Silverman (1994) beschrieben werden. Danach kann der bedingte Erwartungswert der abhängigen Größe Y bei gegebenem Kovariatenset X, U durch

$$E(Y | X, U = u) = X'\beta + g(u)$$

beschrieben werden, wobei g eine hinreichend glatte Funktion in u ist und $X'\beta$ eine lineare Funktion in den übrigen Kovariaten X ist.

U ist hier die regionale Arbeitslosenrate und das Kovariatenset X ist durch die übrigen Merkmale des Humankapitalansatzes gegeben. Die Funktion g ist in dem semiparametrischen Ansatz nicht spezifiziert. Es wird lediglich gefordert, daß g glatt ist in dem Sinne, daß g zweimal stetig differenzierbar ist und

$$\int (g''(u))^2 du$$

erklärt ist. Dieses Integral mißt die Summation der quadrierten Krümmung über den Definitionsbereich von U . Es bestraft damit Funktionen, die stark über den Definitionsbereich von U oszillieren. Die Modellschätzung erfolgt dann auf Basis der Maximierung einer um $\int (g''(u))^2 du$ erweiterten Likelihoodfunktion:

$$\tilde{l}(\beta, g) = l(\beta, g) - \frac{1}{2}\lambda \int (g''(u))^2 du$$

Hierbei ist $l(\beta, g)$ die übliche Loglikelihoodfunktion in den Modellparametern β und g und λ ist ein zu spezifizierender Glättungsparameter.

Es läßt sich zeigen (vgl. Green und Silverman 1994, S. 105), daß sich dieses Maximierungsproblem auf ein Problem in endlich vielen Parametern reduzieren läßt. Hierbei ist $g(u)$ bis auf die Werte $g_i = g(u_i)$ festgelegt. Hierbei sind die u_i die beobachteten Werte der Kovariaten U . Der Funktionsverlauf von $g(u)$ ist dann durch einen natürlichen kubischen Spline durch die Punkte g_i gegeben. Hierbei ist ein kubischer Spline eine zweimal stetig differenzierbare Funktion, die

zwischen den Punkten g_i durch jeweils ein Polynom 3. Grades gegeben ist (vgl. Green und Silverman 1994, S. 11).

Für die Schätzung dieses Modells haben wir das Programmpaket *GAMFIT* (Generalized Additive Model Fitting) von Hastie und Tibshirani benutzt. Die Spezifikation des parametrischen Modellteils stimmt mit der des später vorgestellten Pooled-Modells ohne Regional-Dummies in Tabelle 2 überein. Weiterhin wurde für die Abweichungen von Y vom bedingten Erwartungswert $E(Y | X, U)$ die Normalverteilung unterstellt.

Die Abbildungen 1, 2 und 3 zeigen die geschätzten Punkte¹¹ \hat{g}_i für drei unterschiedliche Werte¹² des Glättungsparameters λ . Zusätzlich sind in den Abbildungen 1 und 2 noch die geschätzten 95 Prozent Konfidenzintervalle für $g(u_i)$ angegeben¹³. Die Abweichungen von der gestrichelten horizontalen Linie sind als prozentuale Abweichungen des Bruttolohns vom Stichprobendurchschnitt bei gegebener Arbeitslosenquote zu interpretieren.

Die Abbildung 1 zeigt, daß man bei starker Glättung (Die Anzahl der verallgemeinerten Freiheitsgrade für g ist 2) einen nahezu perfekten quadratischen Einfluß der regionalen Arbeitslosenquote auf den logarithmierten Bruttolohn erhält. Mit abnehmender Glättung in Abbildung 2 (Anzahl der verallgemeinerten Freiheitsgrade: 8) und Abbildung 3 (Anzahl der verallgemeinerten Freiheitsgrade: 45) zeigt sich jedoch, daß es auch deutliche Abweichungen von einem einfachen quadrierten Zusammenhang zwischen den logarithmierten Bruttolöhnen und der Arbeitslosenquote gibt.

Die hier nicht wiedergegebenen geschätzten Koeffizienten für den linearen Teil des Modells unterscheiden sich nur unwesentlich für die unterschiedlichen Glättungsparameter¹⁴. Weiterhin bleibt die Gestalt des geschätzten Einflusses der Arbeitslosenquote stabil, wenn man für die Merkmale Berufserfahrung und Betriebszugehörigkeitsdauer statt der quadratischen Modellvorgabe ebenfalls einen kubischen Spline benutzt (Die Anzahl der für diese Merkmale benutzten Freiheitsgrade war jeweils 2).

Die Wahl des optimalen Glättungsparameters soll hier nicht diskutiert werden¹⁵, jedoch ist unmittelbar evident, daß:

- a) die günstigste parametrische Spezifikation ein quadratischer Ansatz in der Arbeitslosenquote ist.

¹¹Für die Abbildung 3 wurde aus Gründen der Übersichtlichkeit auf die Wiedergabe der Konfidenzintervalle verzichtet.

¹²In Abbildung 2 wird die geschätzte Splinefunktion $g(u)$ wiedergegeben.

¹³Das Programmpaket *GAMFIT* steuert die Wahl des Glättungsparameters λ über verallgemeinerte Freiheitsgrade für die Störterme, vgl. Green und Silverman 1994, S. 37.

¹⁴Da die geschätzten Koeffizienten nur unwesentlich bei verschiedenen Glättungsparametern variieren, stimmen sie gut mit den Schätzwerten bei der parametrischen Modellspezifikation in Tabelle 2, Spalte 1 überein.

¹⁵Nach dem verallgemeinerten Kreuzvalidierungskriterium von Hastie und Tibshirani 1990, S. 159 ist die in Abbildung 2 wiedergegebene Schätzung optimal.

- b) kein monoton fallender Zusammenhang zwischen der Arbeitslosenquote und dem Erwerbseinkommen erwartet werden kann.

6 Schätzung der Lohngleichung mit gepoolten Individualdaten

Wie die Ergebnisse der semi-parametrischen Analyse gezeigt haben, läßt sich ein nicht-linearer Zusammenhang zwischen Lohnhöhe und Arbeitslosigkeit am ehesten durch die Arbeitslosenquote und deren Quadrat approximieren. Auch wenn dies den bisherigen internationalen und nationalen empirischen Studien widerspricht, in denen sich die Spezifikation mit $\ln(U)$ und $\ln^3(U)$ großer Beliebtheit erfreut, benutzen wir gestützt auf die Ergebnisse des vorherigen Abschnitts in den folgenden parametrischen Modellen ausschließlich die Arbeitslosenquote und deren Quadrat.

Zunächst wird, wie in der Literatur üblich, ein OLS-Modell mit über die Zeitpunkte 1985, 1989 und 1993 gepoolten individuellen Einkommensdaten geschätzt:

$$Y_{i,t} = \ln(W_{i,t}) = \beta_{0,t} + X'_{i,t}\beta + \gamma_1 U_{i,t} + \gamma_2 U_{i,t}^2 + \epsilon_{i,t}$$

$Y_{i,t}$ ist das logarithmierte monatliche Bruttoarbeitseinkommen. X enthält die übrigen lohnbestimmenden Faktoren. Der Index i symbolisiert Personen, der Index t die Zeitpunkte. Im Pooled-Modell wird die Korrelation der $\epsilon_{i,t}$ über die Zeit ignoriert. Die Ergebnisse dieses Modells zeigt die erste Spalte von Tabelle 2. Die geschätzte Humankapitalrendite liegt mit 6,7% je zusätzlichem Ausbildungsjahr im üblichen Rahmen. Das gleiche gilt für die Berufserfahrung und deren Quadrat. Der Lohnzuwachs für jedes zusätzliche Jahr der Betriebszugehörigkeit ist mit einem halben Prozent eher gering, gleichwohl ist auch dieses Ergebnis nicht überraschend. Für drei der fünf in das Modell aufgenommenen Branchen, ergeben sich signifikante Lohndifferenzen zur Referenzgruppe der in der Chemieindustrie Beschäftigten.

Das Modell zeigt einen deutlich signifikanten Zusammenhang zwischen der Lohnhöhe und der Arbeitslosenquote. Gleichwohl ist die Richtung des Zusammenhangs nicht eindeutig, vielmehr zeigt sich wie in der semi-parametrischen Analyse ein parabelförmiger Verlauf.¹⁶ Abbildung 4 zeigt, daß der von Blanchflower und Oswald postulierte monoton fallende Zusammenhang nur für Arbeitslosenquoten bis 8% gelten (die exakte Nullstelle dieser Funktion liegt bei 8,3), danach kehrt sich der Zusammenhang zwischen Lohnhöhe und Arbeitslosigkeit um.

Im negativen Steigungsbereich der Funktion ist der Effekt steigender Arbeitslosigkeit auf den Lohn allerdings nicht sehr hoch (die Elastizitäten des Lohnes

¹⁶Dieses Ergebnis erhält man übrigens auch bei der Modellspezifikation mit $\ln(U)$ und $(\ln(U))^3$.

bezüglich der Arbeitslosigkeit können aus Tabelle 3 entnommen werden). Eine Verdoppelung der Arbeitslosenquote von 4 auf 8% würde nach diesem Modell zu einer durchschnittlichen Lohnsenkung von 4% führen. Dieser Befund stützt also das Lohnkurvenmodell von Blanchflower und Oswald. Nicht erklären läßt sich mit diesem Modell allerdings der Befund eines positiven Zusammenhangs zwischen Arbeitslosigkeit und Lohnhöhe, wenn die Arbeitslosenquote 8% übersteigt. Dieser positive Zusammenhang steht vielmehr im Einklang mit den schon angesprochenen Modellen auf Basis der Theorie der kompensierenden Lohndifferenziale, die gleichwohl auch nicht den gesamten parabelförmigen Verlauf erklären können. Auch die traditionelle neoklassische Theorie legt einen positiven Zusammenhang zwischen Lohnhöhe und Arbeitslosigkeit nahe. Jedoch besteht hier der kausale Zusammenhang nicht zwischen Arbeitslosigkeit und Lohnhöhe, wie von Blanchflower und Oswald postuliert, sondern die Wirkungsrichtung ist genau umgekehrt: Steigende Löhne verringern die Nachfrage nach Arbeit und führen so zu steigender Arbeitslosigkeit. Bevor weiter über diesen Befund spekuliert wird, sollte man sich aber vor Augen halten, daß die Arbeitslosenquote in diesem Eingleichungsmodell auch unbeobachtete Einflüsse regionsspezifischer Besonderheiten widerspiegeln kann.^{17 18}

Im nächsten Schritt wird das Pooled-Regression-Modell deshalb um Dummy-Variablen für die einzelnen Regionen ergänzt, die über die Zeit unveränderliche regionale Merkmale kontrollieren sollen. Im Idealfall sollten dazu für jede der 75 Raumordnungsregionen eine Dummy-Variable in das Modell aufgenommen werden. In der Praxis erweist sich dieses Verfahren allerdings als nicht sehr robust, da einige Raumordnungsregionen nur sehr schwach besetzt sind. In unserem Fall war eine Schätzung mit 75 Dummy-Variablen nicht durchführbar; wir mußten uns auf 65 Variable beschränken. Die Ergebnisse dieser Schätzung finden sich in der zweiten Spalte von Tabelle 2. Die Berücksichtigung von Dummy-Variablen für Raumordnungsregionen führt dazu, daß der Einfluß der Arbeitslosenquote auf die Lohnhöhe völlig verschwindet. Beide Koeffizienten nehmen fast den Wert Null an. Der Vollständigkeit halber ist der Effekt in Abbildung 4 wiedergegeben.

Allerdings ist der Umkehrschluß, daß nämlich nur die unbeobachteten Regressionsmerkmale einen Einfluß auf die Lohnbildung haben, ebenfalls nicht völlig stichhaltig. Da in diesem Modell nur die Abweichungen von den regionsspezifischen Mittelwerten analysiert werden, wird man bei geringen intraregionalen (d.h. zeitlichen) Schwankungen der Arbeitslosenrate die möglicherweise hohen interregionalen Effekte der Arbeitslosenrate nicht erkennen.

¹⁷Die Annahme, daß unbeobachtete regionsspezifische Merkmale mit der Arbeitslosenrate korrelieren, bedeutet eine Verletzung der Modellannahmen des Pooled Modells.

¹⁸Letztendlich läßt sich die Bestimmung eines Kausaleffekts nicht im Rahmen eines Eingleichungsmodells entscheiden. Dies wäre erst im Rahmen eines Mehrgleichungsmodells behandelbar. Die notwendigen Annahmen zur Identifikation eines derartigen Modells wären allerdings kritisch zu hinterfragen.

7 Die Kontrolle regionaler Effekte durch ein verallgemeinertes Varianzkomponentenmodell

Dies wirft die Frage auf, ob nicht noch weitere Möglichkeiten bestehen, nicht beobachtete regionsspezifische Einflüsse zu kontrollieren. Eine Möglichkeit besteht darin, die in individuellen Paneldaten enthaltene Information in einem verallgemeinerten Varianzkomponenten-Modell (GEC) zu nutzen. In der Einkommensanalyse werden Paneldaten im allgemeinen genutzt, um die unbeobachtete Heterogenität der Personen mit einer Varianzkomponente zu identifizieren. Wir führen in unserem Modell zusätzlich eine regionsspezifische Varianzkomponente ein, um tatsächlich auf die Regionen zurückgehende Unterschiede identifizieren zu können. Das Modell hat dann die folgende Form:

$$Y_{i,r,t} = \beta_{0,t} + X'_{i,r,t}\beta_X + \epsilon_{i,r,t}$$

mit i als Personenindex, r als Index für die Region und t als Zeitindex. Der Störterm wird in folgende Varianzkomponenten zerlegt:

$$\epsilon_{i,r,t} = \delta_i + \delta_{i,r} + \delta_{i,r,t}$$

Die durch $\delta_{i,r}$ repräsentierte Varianzkomponente mißt die Interaktionseffekte der Arbeitnehmer mit der Region. Dies sind also diejenigen Einkommensschwankungen, die durch Regionswechsel der Arbeitnehmer entstehen. Um $\delta_{i,r}$ identifizieren zu können, müssen natürlich Regionswechsel von Arbeitnehmern in der Stichprobe beobachtet werden. Mit zunehmender Dauer des SOEP wird das Auftreten derartiger Regionswechsel immer wahrscheinlicher.

Unter dem Modellaspekt wäre die Aufnahme einer reinen Regionskomponente δ_r mit

$$\epsilon_{i,r,t} = \delta_i + \delta_r + \delta_{i,r} + \delta_{i,r,t}$$

wünschenswert. Da für jede Region auch Beobachtungen von mehreren Arbeitnehmern in der Stichprobe sind, ist eine derartige Komponente auch identifiziert. Allerdings haben die dabei auftretenden Kovarianzmatrizen keine blockdiagonale Struktur, so daß bei einem derartigen Modell die im folgenden benutzten Resultate nicht mehr benutzt werden können.

Wir machen im folgenden die Annahme, daß die 3 Varianzkomponenten unabhängig normalverteilt sind:

$$\delta_i \sim N(0, \sigma_I^2)$$

$$\delta_{i,r} \sim N(0, \sigma_R^2)$$

$$\delta_{i,r,t} \sim N(0, \sigma_T^2)$$

Damit hat die Kovarianzmatrix Σ_i der Störterme einer Person i die folgende blockdiagonale Gestalt:

$$\epsilon'_{i,r} = (\epsilon_{i,r,t_1}, \dots, \epsilon_{i,r,t_{n_{i,r}}})' \quad n_{i,r} = \text{Anzahl Beob. von Pers. } i \text{ in Region } r$$

$$\epsilon'_i = (\epsilon'_{i,r_1}, \dots, \epsilon'_{i,r_{n_i}}) \quad n_i = \text{Anzahl Regionen für Person } i$$

$$\begin{aligned} \text{Cov}(\epsilon_i) = \Sigma_i &= \sigma_T^2 \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} + \sigma_I^2 \begin{pmatrix} 1 & \dots & 1 \\ \vdots & & \vdots \\ 1 & \dots & 1 \end{pmatrix} \\ &+ \sigma_R^2 \begin{pmatrix} 1 & \dots & 1 & & & \\ \vdots & & \vdots & & & 0 \\ 1 & \dots & 1 & & & \\ & & & \ddots & & \\ & & & & 1 & \dots & 1 \\ & & 0 & & \vdots & & \vdots \\ & & & & 1 & \dots & 1 \end{pmatrix} \end{aligned}$$

Die blockdiagonale Struktur der Gesamtmatrix aller Störterme ermöglicht deren explizite Inversion. Unter Zuhilfenahme eines allgemeinen Resultats von Sakalauskas (1993) erhält man für den Likelihoodbetrag einer Person:

$$\begin{aligned} -2\ln L_i &= m_i \ln(2\pi) + \ln(\det(\Sigma_i)) + \epsilon'_i \Sigma_i^{-1} \epsilon_i \\ &= m_i \ln(2\pi) + \ln \left[\left(\prod_r n_{i,r} (\sigma_R^2 + \sigma_T^2/n_{i,r}) (\sigma_T^2)^{n_{i,r}-1} \right) \right. \\ &\quad \left. \times \left(1 + \sigma_I^2 \sum_r \frac{1}{\sigma_R^2 + \sigma_T^2/n_{i,r}} \right) \right] \\ &+ \left\{ \frac{1}{\sigma_T^2} \sum_i (e_{i,r,t} - e_{i,r,\bullet})^2 \right. \\ &\quad + \sum_r \frac{1}{\sigma_R^2 + \sigma_T^2/n_{i,r}} (e_{i,r,\bullet} - e_{i,\bullet,\bullet})^2 \\ &\quad \left. + \frac{\sum_r \frac{1}{\sigma_R^2 + \sigma_T^2/n_{i,r}}}{\sigma_I^2 (\sum_r \frac{1}{\sigma_R^2 + \sigma_T^2/n_{i,r}}) + 1} e_{i,\bullet,\bullet}^2 \right\} \end{aligned}$$

wobei m_i = Anzahl aller Beobachtungen für Person i und:

$$\begin{aligned}
 e_{i,r,t} &= Y_{i,r,t} - \beta_{0,t} - X'_{i,r,t} \beta_X \\
 e_{i,r,\bullet} &= \frac{1}{n_{i,r}} \sum_t e_{i,r,t} \\
 e_{i,\bullet,\bullet} &= \frac{1}{\sum_r \frac{1}{\sigma_R^2 + \sigma_T^2/n_{i,r}}} \sum_r \frac{1}{\sigma_R^2 + \sigma_T^2/n_{i,r}} e_{i,r,\bullet}
 \end{aligned}$$

- Für Personen mit nur einer Region hängt $\ln L_i$ nur über $\sigma_I^2 + \sigma_R^2$ von σ_I^2 und σ_R^2 ab.
- Für Personen mit nur einer Beobachtung pro Region hängt $\ln L_i$ nur über $\sigma_R^2 + \sigma_T^2$ von σ_R^2 und σ_T^2 ab.

Für die folgenden Schätzungen (Spalte 4 in Tabelle 2) wurde die ML-Lösung ermittelt. Als Startwert für die iterative Bestimmung dieser Lösung wurden die Ergebnisse für das Standard Varianzkomponenten Modell mit $\sigma_F = 0$ benutzt. Auch für dieses Standardmodell wurde die ML Lösung bestimmt. Diese Werte sind in Spalte 3 von Tabelle 1 ausgewiesen.

Im Regelfall sollten die Regionen, auf deren Basis die Arbeitslosenquoten in das Modell eingehen und die Regionen, die die Varianzkomponente identifizieren, identisch sein. Dies würde erfordern, daß man eine ausreichende Zahl von Personen hat, die im betrachteten Zeitraum mindestens einmal die Raumordnungsregion gewechselt haben. In dem hier verwendeten Sample trifft das allerdings nur auf 22 Personen zu und diese Fallzahl ist für die Identifizierung von $\delta_{i,r}$ sicherlich nicht ausreichend. Für die hier präsentierte Analyse verwenden wir für Regionswechsel deshalb nicht die Raumordnungsregionen, sondern die Kreise bzw. kreisfreien Städte. Die Kreiskennziffer des Wohnorts ist für jede Person in jedem Jahr bekannt. Anhand der Veränderungen können Regionswechsler identifiziert werden¹⁹. Tabelle 4 gibt einen Überblick zu den in dem hier verwendeten Sample auftretenden Mustern von Regionswechseln.

Die Tabelle 4 zeigt, daß 94.2% der Stichprobenmitglieder an den drei Referenzjahren stets innerhalb desselben Kreises wohnen. Diese Quote ist für einen Zeitraum von fast 10 Jahren relativ hoch und auf jeden Fall größer als es die durchschnittliche jährliche Umzugsrate von 5% aller Haushalte erwarten läßt. Dies erklärt sich zum einen dadurch, daß das Gros der Umzüge die Kreisgrenzen nicht überschreitet. Zum anderen werden bestimmte Regionswechsler systematisch nicht erfaßt, weil die betreffenden Arbeitnehmer vor dem Wechsel entweder noch nicht voll erwerbstätig waren oder weil sie nach dem Wechsel nicht mehr erwerbstätig sind. Auch die Panelmortalität, bei der verstärkt Personen mit

¹⁹Aus Datenschutzgründen sind diese Informationen jedoch nicht allgemein zugänglich und können nur am DIW genutzt werden.

Umzügen die Stichprobe verlassen (Rendtel 1994, Kapitel 8), kann zu diesem geringen Anteil von beobachteten Regionalwechslern beigetragen haben.

Zunächst wollen wir die Ergebnisse eines einfachen Varianzkomponenten-Modells diskutieren, in dem nur eine individuenspezifische Störgröße berücksichtigt wird (EC-Modell in Spalte 3 von Tabelle 1). In der Literatur über Lohnkurvenschätzung ist diese Modellvariante bislang noch nicht zu finden. Im Vergleich mit dem einfachen Pooled-Regression-Modell geht der Einfluß der Arbeitslosigkeitsrate auf die Lohnhöhe deutlich zurück. Der Koeffizient der Arbeitslosenquote steigt von -0.05 auf -0.019 und der Koeffizient des quadrierten Terms sinkt von 0.003 auf 0.001. Beide Koeffizienten sind zudem nur noch auf einem sehr niedrigen Niveau signifikant. Abbildung 4 zeigt, daß auch diese Lohnkurve einen parabelförmigen Verlauf hat, allerdings liegt die Nullstelle der Funktion nicht bei einer Arbeitslosenquote von 8.3, wie im Pooled-Modell, sondern bei 9.5. Die geschätzten Einkommenseffekte sind in diesem Modell deutlich geringer (die Elastizitäten sind Tabelle 3 zu entnehmen). Eine Verdoppelung der Arbeitslosenquote von 4 auf 8% führt zu einem geschätzten Einkommensrückgang von etwa 2%.

Festgehalten werden sollte schon an dieser Stelle, daß sich der Zusammenhang zwischen Lohnhöhe und Arbeitslosigkeit schon dann wesentlich abschwächt, wenn statt eines Pooled-Regression-Modells ein einfaches Varianzkomponenten-Modell genutzt wird, das die individuelle unbeobachtete Heterogenität der Beschäftigten kontrolliert. Es sei aber daran erinnert, daß für das Gros der Stichprobe die Region über die Zeit konstant ist und daß folglich im EC Modell die Varianzkomponente δ_i den Einfluß aller zeitlich konstanten personenbezogenen Merkmale auf den Lohn mißt, also auch den Einfluß der Region. Aus diesem Grund ist die in Abbildung 4 dokumentierte Zwischenstellung des EC Modells zwischen dem Pooled Modell mit Regional Dummies durchaus plausibel.

In der 4. Spalte von Tabelle 1 sind schließlich die Ergebnisse des GEC-Modells wiedergegeben, in dem mittels einer spezifischen Varianzkomponente Interaktionen von Personen und Regionen kontrolliert werden. Die Schätzwerte für die Parameter stimmen im wesentlichen mit denen des EC-Modells überein. Die Tatsache, daß keine größeren Abweichungen zum EC-Modell auftreten, dürfte vor allem an der doch sehr geringen Anzahl von Regionswechselern liegen. Insbesondere sind die geschätzten Parameter der Arbeitslosenquote und deren Quadrat in beiden Modellen identisch.

Angesichts des geringen geschätzten Effekts der regionalen Arbeitslosenquote und seiner niedrigen Signifikanz könnte man zu dem Schluß kommen, daß Regionalmerkmale in Deutschland einen vernachlässigbaren Einfluß auf die Lohnbildung haben. Ein derartiger Schluß würde auch durch den den theoretischen Vorhersagen widersprechenden nicht monotonen Einfluß der Arbeitslosenrate auf den Lohn nahegelegt.

Trotzdem ist ein derartiger Schluß ungerechtfertigt. Im unteren Teil von Tabelle 2 sind die geschätzten Varianzen des GEC Modells aufgelistet. Hierbei ist auch die Varianzkomponente σ_R^2 signifikant von 0 verschieden. Aus-

gekräftiger als die Varianzen sind allerdings die geschätzten Standardabweichungen $\hat{\sigma}_T = 0.15$, $\hat{\sigma}_I = 0.21$ und $\hat{\sigma}_R = 0.09$. Das GEC Modell postuliert damit sehr wohl die Existenz von Regionaleffekten in einer durchaus bemerkenswerten Größe. Allerdings erscheint nach den von uns präsentierten Ergebnissen die regionale Arbeitslosenrate ein ungeeigneter Indikator für diese Lohneffekte zu sein.

Insgesamt stehen die von uns präsentierten Resultate unter einigen Vorbehalten:

- a) Es wäre wünschenswert, daß die Regionalindikatoren für Arbeitslosigkeit und Regionalwechsel identisch sind.
- b) Aufgrund der geringen Fallzahl der Regionalwechseler sind die Schätzungen für σ_I^2 und σ_R^2 hoch miteinander korreliert ($r=-0.85$).

Beide Einwände würden für einen Datensatz mit deutlich höherer Fallzahl als beim SOEP — beispielsweise die Beschäftigtenstatistik²⁰ — nicht gelten und es wäre daher sinnvoll, das hier vorgestellte Modell mit einem derartigen Datensatz zu schätzen.

Der dritte Einwand ist methodischer Natur. Es ist wünschenswert, das GEC Modell um eine reine Regionalkomponente δ_r zu erweitern. In diesem Fall sind allerdings die Störterme unterschiedlicher Personen nicht mehr unabhängig. Schlimmer noch: Die Zeitpunkte, für die verschiedene Personen über die Komponente δ_r miteinander verknüpft sind, ändert sich von Zeitpunkt zu Zeitpunkt. Trotzdem haben die auftretenden Kovarianzmatrizen noch eine verhältnismäßig einfache Struktur, so daß deren algebraische Invertierung nicht aussichtslos erscheint.

²⁰Voraussetzung ist freilich, daß die Beschäftigtenstatistik regionalisierbar ist und als Panel ausgewertet wird.

8 Tabellen und Abbildungen

Tabelle 1: Mittelwerte und Standardabweichungen (in Klammern) der verwendeten Merkmale.

	1985	1989	1993
Ausbildung in Jahren	11.292 (1.991)	11.350 (2.111)	11.458 (2.200)
Berufserfahrung in Jahren	21.041 (10.352)	22.999 (11.122)	24.927 (10.279)
Berufszugehörigkeit in Jahren	9.615 (7.774)	10.586 (8.480)	11.848 (8.521)
Arbeitszeit Std./Woche	43.813 (9.695)	43.782 (7.944)	43.241 (7.850)
Arbeiter (Dummy)	0.575	0.550	0.541
Branchen (Dummies):			
-KFZ, Eisen, Stahl	0.272	0.287	0.281
-Nahrung, Textil, Holz	0.186	0.210	0.221
-Baugewerbe	0.121	0.117	0.118
-Handel	0.030	0.032	0.039
-Sonstige Dienstleistungen	0.102	0.111	0.104
Arbeitslosenquote in %	7.032 (2.319)	7.199 (2.574)	8.219 (2.143)
Beobachtungen	896	1088	847
Quelle: SOEP			

Tabelle 2: Lohnhöhe und Arbeitslosigkeit, Ergebnisse der Regressionsanalysen 1985 bis 1993 (Standardabweichungen in Klammern).

Variable	Pooled-Model	Pooled-Model mit 65 Reg.-Dummies	Error Component	GEC
Konstante	6.782 (72.877)	6.586 (49.946)	6.736 (76.810)	6.746 (76.896)
1989	0.128 (10.620)	0.132 (11.101)	0.134 (17.824)	0.135 (17.925)
1993	0.322 (24.472)	0.322 (20.282)	0.333 (34.112)	0.333 (34.107)
Ausbildung in Jahren	0.067 (23.655)	0.065 (22.924)	0.076 (19.465)	0.076 (19.447)
Berufserfahrung	0.029 (12.476)	0.028 (12.055)	0.033 (14.763)	0.033 (14.659)
Berufserfahrung ² /100	-0.049 (-10.346)	-0.048 (-10.011)	-0.055 (-12.071)	-0.054 (-11.966)
Betriebszugehörigkeit	0.005 (2.448)	0.004 (2.111)	0.003 (1.711)	0.003 (1.616)
Betriebszugehörigkeit ² /100	-0.005 (-0.741)	-0.002 (-0.260)	-0.004 (-0.577)	-0.004 (-0.519)
ln (Arbeitszeit)	0.129 (7.190)	0.129 (7.286)	0.060 (4.489)	0.059 (4.409)
Arbeiter	-0.227 (-18.723)	-0.219 (-18.149)	-0.148 (-11.116)	-0.149 (-11.208)
Branchen (Ref. Chemie)				
-KFZ, Eisen, Stahl	0.026 (1.865)	0.016 (1.116)	0.003 (0.247)	0.002 (0.215)
-Nahrung, Textil, Holz	-0.005 (-0.308)	0.001 (0.052)	0.010 (0.803)	0.011 (0.815)
-Baugewerbe	-0.013 (-0.708)	-0.017 (-0.958)	-0.021 (-1.191)	-0.020 (-1.163)
-Handel	-0.151 (-5.153)	-0.138 (-4.765)	-0.042 (-1.834)	-0.043 (-1.877)

Tabelle 2: Fortsetzung

Variable	Pooled- Model	Pooled- Model mit 65 Reg.- Dummies	Error Component	GEC
-Sonstige Dienstleistungen	-0.031 (-1.690)	-0.029 (-1.590)	-0.024 (-1.332)	-0.024 (-1.294)
Arbeitslose in %	-0.050 (-3.864)	-0.002 (-0.070)	-0.019 (-1.569)	-0.019 (-1.560)
(Arbeitslose in %)* * 2	0.003 (3.927)	0.000 (0.150)	0.001 (1.660)	0.001 (1.570)
σ_{ϵ}^2	0.070	-	-	-
σ_T^2	-	-	-	-
σ_I^2	-	-	0.023 (28.653)	0.023 (28.115)
σ_R^2	-	-	0.050 (19.153)	0.043 (8.504)
	-	-	-	0.008 (1.736)
Anzahl der Beobachtungen: n = 2831				
Quelle: SOEP 1985, 1989, 1993.				

Tabelle 3: Elastizitäten des Lohnes bezüglich der Arbeitslosigkeit.

Arbeitslosenquote in Prozent	Elastizität	
	Pooled-Model ohne Reg. Dummies	EC/GEC-Model
4	-0.013	-0.0053
5	-0.012	-0.0054
6	-0.010	-0.0051
7	-0.007	-0.0043
8	-0.002	-0.0029
9	0.004	-0.0011
10	0.012	-0.0012
11	0.021	0.004
12	0.032	0.0073
13	0.044	0.011
14	0.057	0.015

Tabelle 4: Häufigkeit von Regionswechseln (Wechsel der Kreiskennziffer) in der Stichprobe.

Anzahl der Beobachtungen in			Häufigkeit	Anteil
1.Region	2.Region	3.Region		
2	-	-	504	43.4
3	-	-	589	50.8
1	1	-	27	2.3
1	2	-	18	1.6
2	1	-	17	1.5
1	1	1	5	0.4

Abbildung 1: Geschätzter Einfluß der regionalen Arbeitslosenrate auf das logarithmierte Einkommen von Männern (Vollzeiterwerbstätige Männer ohne Beamte und Angestellte im öffentlichen Dienst). Schätzergebnisse für ein semi-parametrisches Regressionsmodell. Lineare Koeffizienten wie in Tabelle 2, Spalte 1. Glättungsparameter: 2 Freiheitsgrade für die Residuen (starke Glättung). Darstellung: 95% Konfidenzintervalle um die geschätzten Werte $g(u_i)$.

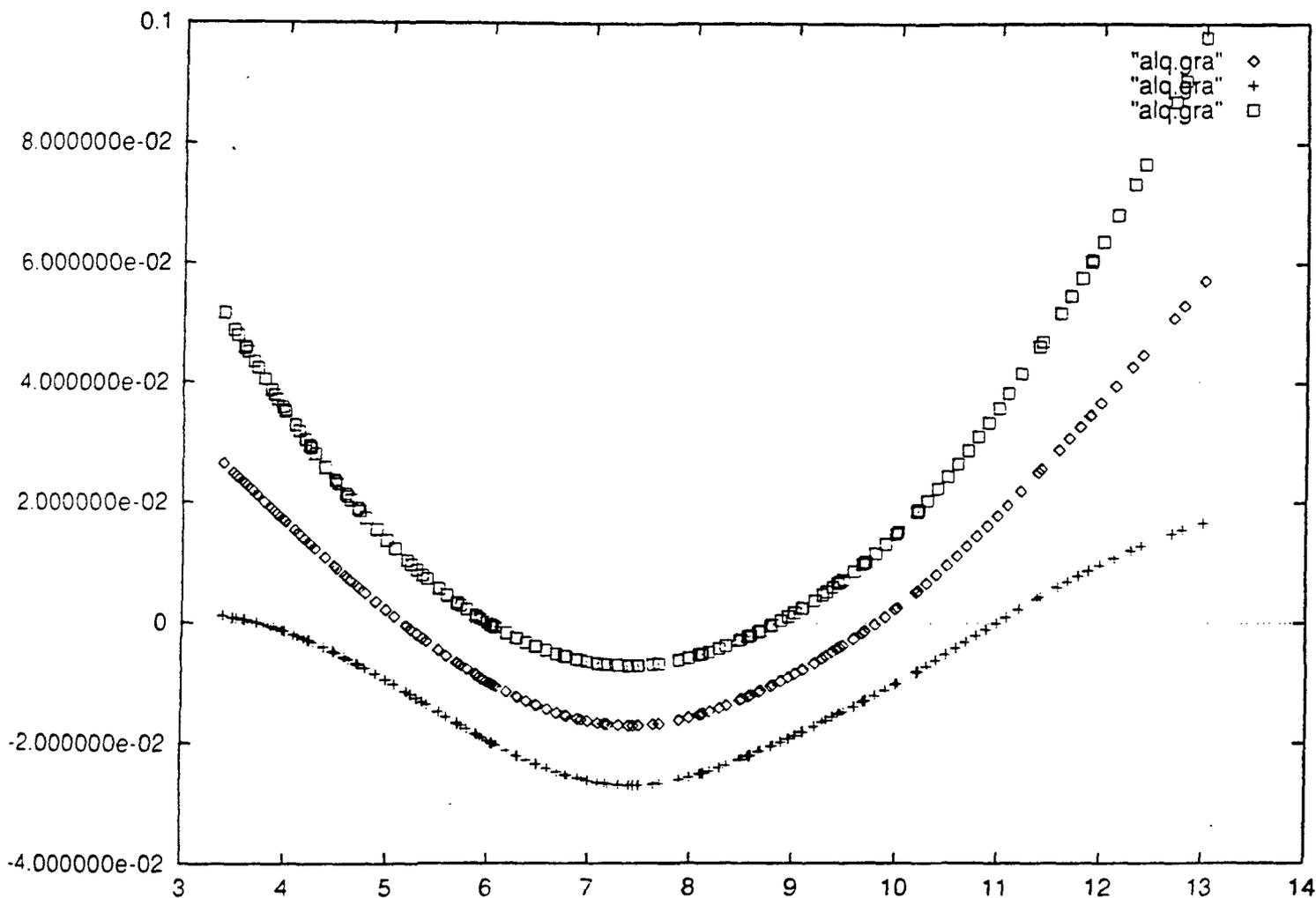


Abbildung 2: Geschätzter Einfluß der regionalen Arbeitslosenrate auf das logarithmierte Einkommen von Männern (Vollzeiterwerbstätige Männer ohne Beamte und Angestellte im öffentlichen Dienst). Schätzergebnisse für ein semi-parametrisches Regressionsmodell. Lineare Koeffizienten wie in Tabelle 2, Spalte 1. Glättungsparameter: 8 Freiheitsgrade für die Residuen (mittlere Glättung). Darstellung: 95% Konfidenzintervalle um die geschätzten Werte $g(u)$.

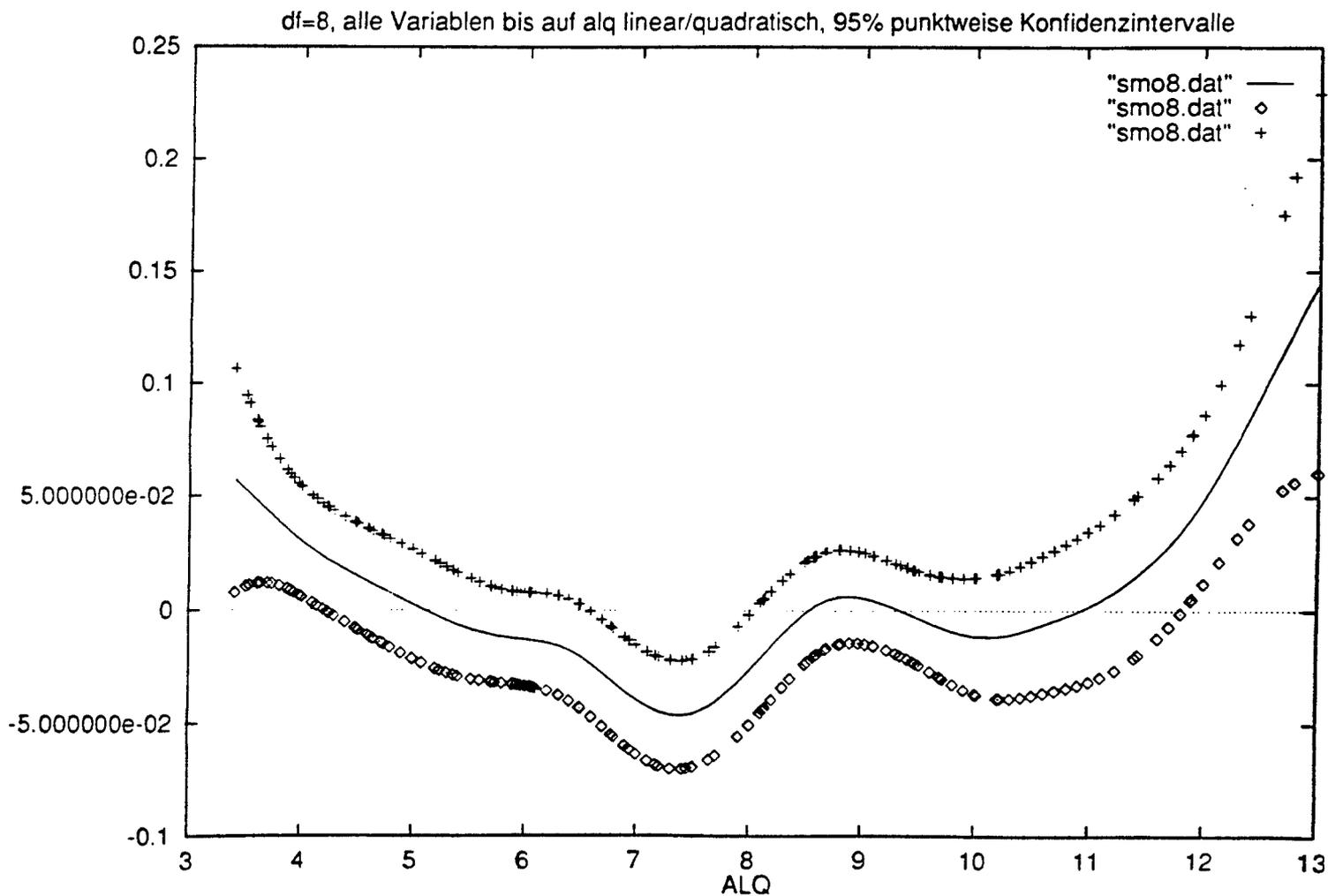


Abbildung 3: Geschätzter Einfluß der regionalen Arbeitslosenrate auf das logarithmierte Einkommen von Männern (Vollzeiterwerbstätige Männer ohne Beamte und Angestellte im öffentlichen Dienst). Schätzergebnisse für ein semi-parametrisches Regressionsmodell. Lineare Koeffizienten wie in Tabelle 2, Spalte 1. Glättungsparameter: 45 Freiheitsgrade für die Residuen (sehr schwache Glättung). Darstellung: Geschätzte Werte $g(u_i)$.

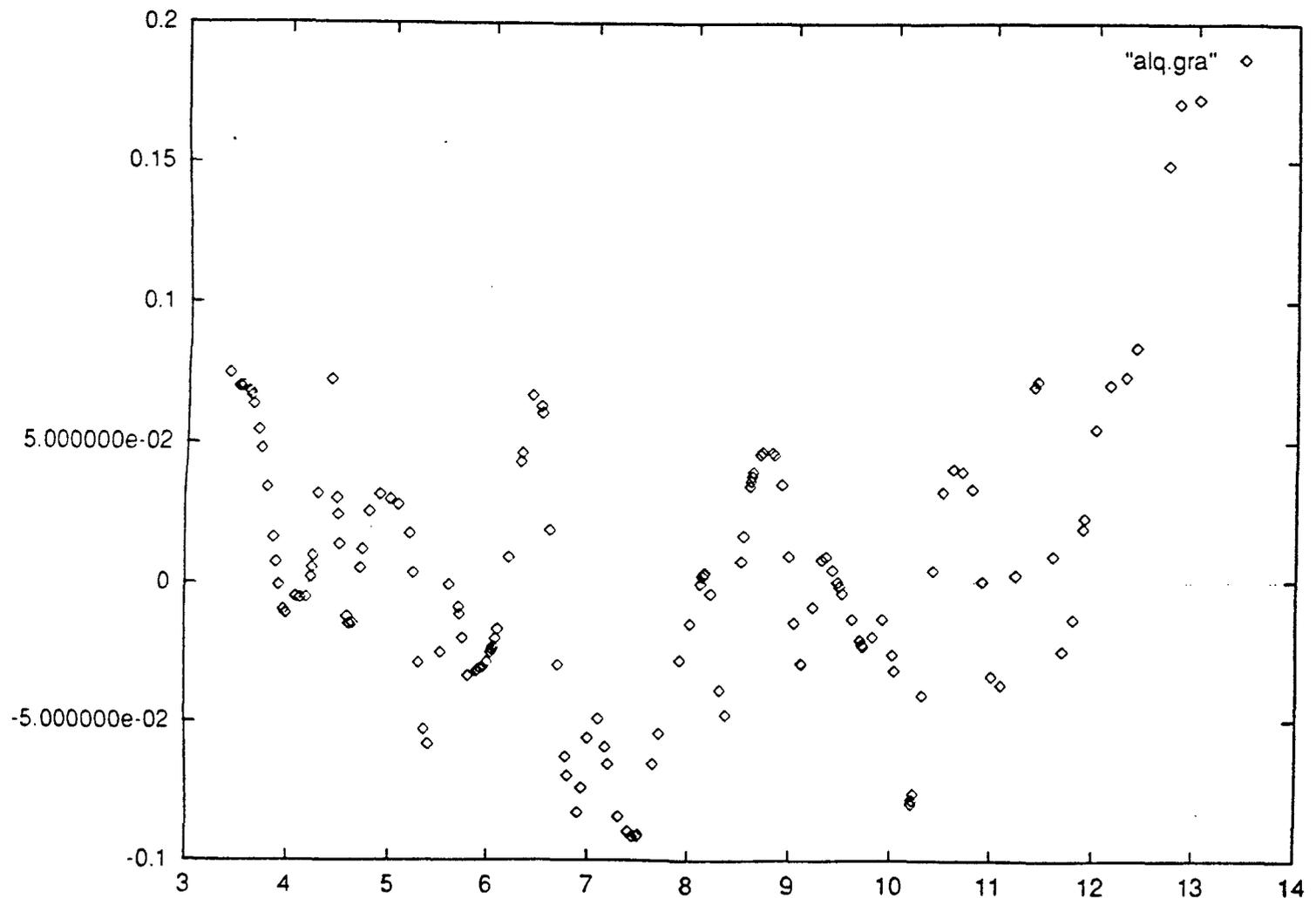
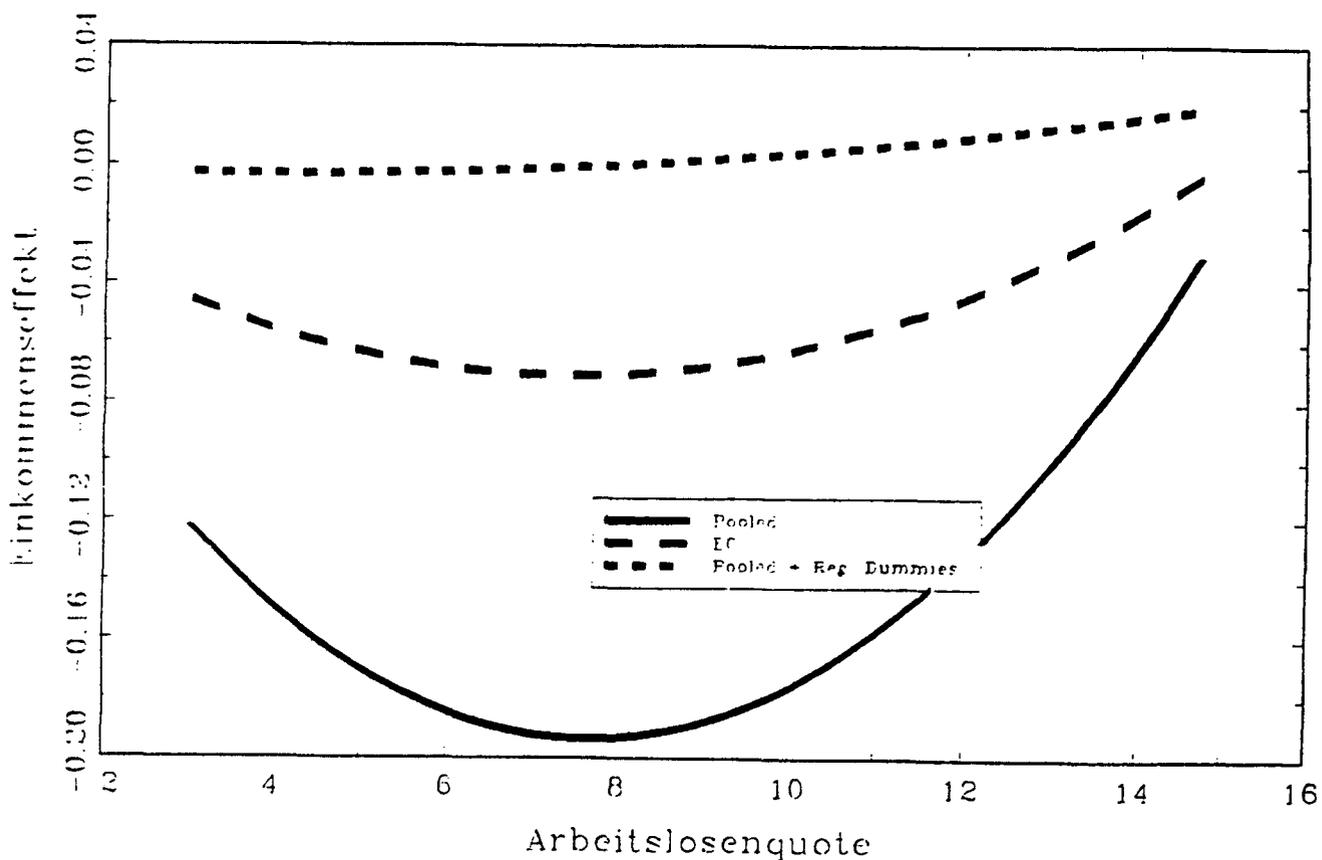


Abbildung 4: Geschätzter Einfluß der regionalen Arbeitslosenrate auf das logarithmierte Einkommen von Männern (Vollzeiterwerbstätige Männer ohne Beamte und Angestellte im öffentlichen Dienst). Schätzergebnisse für drei parametrische Modelle. Berücksichtigung der Arbeitslosenrate durch einen quadratischen Ansatz.

Geschaetzter Einfluss der Arbeitslosenrate
Modell: $u + u^2$



Literatur

- Blanchflower, David G. und Andrew J. Oswald 1990:* The Wage Curve. In: *Scandinavian Journal of Economics*, (92), S. 215-235.
- Blanchflower, David G. und Andrew J. Oswald (1992):* International Wage Curves. NBER Working Paper, No. 4200, Cambridge.
- Blanchflower, David G. und Andrew J. Oswald 1994:* Estimating a Wage Curve for Britain 1973-90. In: *The Economic Journal*, (104), S. 1025-1043.
- Bundesforschungsanstalt für Landeskunde und Raumordnung (BfLR) 1992:* Laufende Raumb Beobachtung. Aktuelle Daten zur Entwicklung der Städte, Kreise und Gemeinden 1989/90. Materialien zur Raumentwicklung Heft 47. Bonn.
- Gerlach, Knut und Joachim Wagner 1994:* Regionale Lohnunterschiede und Arbeitslosigkeit in Deutschland. Ein Beitrag zur Lohnkurven-Diskussion. Universität Lüneburg, Fachbereich Wirtschafts- und Sozialwissenschaften, Arbeitsbericht Nr. 131. Lüneburg.
- Groot, Wim, Eddie Meekelholt und Hessel Oosterbeek 1992:* Further Evidence on the wage curve. In: *Economics Letters*, (38), S. 355-359.
- Gornig, Martin und Johannes Schwarze 1992:* Möglichkeiten und Grenzen regionaler Einkommensanalysen mit dem Sozio- ökonomischen Panel. Universität Bremen, ZWE Arbeit und Region, Arbeitspapier Nr. 1. Bremen.
- Green, P.J.; Silverman, B.W. 1994:* Nonparametric and Generalized Linear Models. A roughness Penalty Approach. Chapman and Hall, London.
- Hall, Robert E. 1972:* Turnover in the Labor Force. In: *Brookings Papers on Economic Activity*, (3), S. 709-764. Harris, John R. und Michael P. Todaro (1970): Migration, Unemployment and Development. In: *American Economic Review*, (60), S. 126-142.
- Hastie, T.,J.; Tibshirhani, R.J. 1990:* Generalized Additive Models. Chapman and Hall, London.
- Hsiao, Cheng 1995:* Panel Analysis for Metric Data. In: G. Arminger, C. C. Clogg und M. E. Sobel (Hg.): *Handbook of Statistical Modeling for the Social and Behavioral Sciences*. New York. S. 361-400.
- Jacoby, Herbert 1990:* Regionale Lohnstrukturen. Eine theoretische und empirische Analyse regionaler Lohnunterschiede. Frankfurt/New York.
- Moulton, R. B. 1990:* An illustration of a pitfall in estimating the effects of aggregate variables on micro units. In: *Review of Economics and Statistics*, (72), S. 334-338.
- Projektgruppe Sozio-oekonomisches Panel 1995:* Das Sozio- oekonomische Panel (SOEP) im Jahre 1994. In: *Vierteljahrshefte zur Wirtschaftsforschung*, (64), S. 5-15.
- Rendtel, Ulrich 1994:* Die Analyse von Paneldaten unter Berücksichtigung der Panelmortalität. Theorie und Empirie am Beispiel des Sozio-oekonomischen Panels (SOEP), Deutsches Institut für Wirtschaftsforschung, Berlin (Erscheint 1995 im Campus Verlag).
- Sakalauskas, L. 1993:* Application of Matrices with Constant Blocks in a High-Dimensional Statistical Analysis. In: Janssen, J.; Skiadas, C. (eds): *Applied Stochastic Models and Data Analysis*, World Scientific Publ., New Jersey, 815-823.
- Topel, Robert 1986:* Local Labor Markets. In: *Journal of Political Economy*, (94), S. S111-S143.
- Wagner, Joachim 1994a:* German wage curves, 1979-1990. In: *Economics Letters*, (44), S. 307-311.
- Wagner, Joachim 1994b:* Arbeitslosigkeit und regionale Lohndifferentiale: Lohnkurven für Niedersachsen 1979 - 1990. Universität Lüneburg, Fachbereich Wirtschafts- und Sozialwissenschaften, Arbeitsbericht Nr. 137. Lüneburg.