Neuere Ansätze in der Mikroökonometrie

Modellierung, Schätz- und Testverfahren

Olaf Hübler*

Diskussionspapier Nr. 270
Februar 2003
ISSN 0949-9962

Anschrift:
Universität Hannover
Institut für Quantitative Wirtschaftsforschung
Königsworther Platz 1
30167 Hannover

Fax: 0511-762-3923
Email: huebler@mbox.iqw.uni-hannover.de

* Bernd Fitzenberger, Norbert Janz, Gerd Ronning sowie den Teilnehmern des 32. Wirtschaftswissenschaftlichen Seminars Ottobeuren danke ich für hilfreiche Kommentare

1 Einführung

Die erste Phase in der Ökonometrie von 1930-1970 war sowohl im methodischen Bereich als auch bei den Anwendungen eindeutig geprägt von makroökonomischen Fragestellungen. Gesamtwirtschaftliche Ansätze keynesianischer und neoklassischer Provenienz standen bei der Modellierung Pate. Dies war nicht nur durch die vorherrschende Vorgehensweise in der Volkswirtschaftslehre der damaligen Zeit, sondern auch durch die nahezu ausschließliche Verfügbarkeit aggregierter Daten bedingt. Ab Mitte der siebziger Jahre lässt sich eine Umorientierung beobachten, verbunden mit Entwicklungen, die die Entstehung der Mikroökonometrie begünstigten. Versagen der Wirtschaftspolitik bei der Lösung der Probleme, die durch die beiden Ölkrisen ausgelöst wurden, nachlassendes Vertrauen in die Wirksamkeit globaler Maßnahmen und Lucas-Kritik haben zur Hinwendung zu mikroökonomischen Ansätze beigetragen. Parallel dazu verlief die Entwicklung leistungsfähiger Computer, die die Verarbeitung großer Datenmengen zuließ. Zwar wurden mikroökonomische Daten bereits früher erhoben. Sie standen allerdings der Wissenschaft in Form von Einzeldaten nicht zur Verfügung, sondern wurden von den statistischen Ämtern aggregiert und nur so bereitgestellt. Initiativen der Wissenschaft, eigene Datensätze zu erheben, zunächst auf Individuen und Haushalte beschränkt, später ausgedehnt auf Betriebe, haben den Informationsstand über einzelwirtschaftliche Phänomene deutlich verbessert und die Verfügbarkeit der entsprechenden Daten möglich gemacht. In der Zwischenzeit ist auch die amtliche Statistik aufgeschlossener Dritten gegenüber, diese mit aus der Primärstatistik gewonnenen Mikrodaten arbeiten zu lassen.

Während mikroökonomische Fragestellungen unter Verwendung der entsprechenden Daten ökonometrisch zunächst überwiegend mit den gleichen Methoden bearbeitet wurden wie gesamtwirtschaftliche Ansätze, führten offensichtlich zu Tage tretende Defizite sehr schnell zu Weiterentwicklungen. Bedingt durch die theoretische Herangehensweise und die Art der Daten wurde es notwendig, neue Methoden zu entwickeln. Heterogene Verhaltensweisen der Wirtschaftssubjekte, die auch Änderungen unterliegen, erhalten einen viel größeren Stellenwert als in der traditionellen, gesamtwirtschaftlich orientierten Ökonometrie. Kausale Zusammenhänge treten stärker in den Vordergrund. Gruppenspezifisches Verhalten muss genauso modelliert werden wie die Tatsache, dass es unbeobachtete Einflüsse gibt. Zudem ist die Qualität der Daten eine andere. Gesamtwirtschaftlich war das Interesse fast ausschließlich auf quantitative, kardinal gemessene Informationen gerichtet. Hierauf kann und will sich die empirische Analyse mikroökonomischer Phänomene nicht beschränken. Erstens lassen sich Verhaltensweisen und Entscheidungen nicht auf einer kontinuierlichen Skala abbilden. Und zweitens, selbst wenn dies prinzipiell möglich ist, sind Wirtschaftssubjekte nicht in jedem Fall bereit, detaillierte Auskünfte zu geben. Ja-Nein-Angaben oder abgestufte Angaben wie "besser, schlechter, gleich" sind keine Seltenheit. Drittens können die möglichen Werte auch Beschränkungen unterliegen. Es lassen sich zum Teil nur ganze nichtnegative Zahlen realisieren, wie z.B. bei der Zahl der Arbeitsplatzwechsel, oder der Wertebereich ist nach oben bzw. unten beschränkt. Ein negatives Arbeitsangebot z.B. ist nicht möglich. Die Zeit, die ein bestimmter Zustand andauert, eine Arbeitslosigkeitsperiode, ist bei der Erhebung u.U. noch nicht abgeschlossen. Oder es fehlen Informationen darüber, wann z.B. ein Zustand eingetreten ist, wann ein Betrieb Gewinnbeteiligung eingeführt hat. Aus methodischer Sicht hat sich aufgrund dieser Tatbestände der Kern der Mikroökonometrie (Ronning 1991) entwickelt:

Modelle mit qualitativen Variablen

Modelle mit ordinalen Daten

Modelle mit Zähldaten

Modelle mit zensierten und abgeschnittenen Variablen

Selektionsmodelle

Zeitdauermodelle.

Üblicherweise konzentriert sich die Methodik, der es vorrangig um die Spezifikation, das Schätzen von Parametern und das Testen von Hypothesen geht, auf die endogene Variable. Nicht kardinalgemessenen exogenen Variablen wird dagegen kaum Aufmerksamkeit geschenkt. Abgesehen von der Qualität und Messung der Daten spielen die Periodizität, die Art der Erhebung sowie der Datenträger eine Rolle. Hiernach lässt sich eine andere Einteilung der Mikroökonometrie vornehmen:

Analysen mit Querschnittsdaten

Analysen mit Paneldaten

Analysen der Verweildauer

Analysen aus kombinierten Individual- und Betriebsdaten.

Analysen von Hochfrequenzdaten

In der ersten Zeit war die Mikroökonometrie, bedingt durch die Datenlage, vollständig mit Querschnittsanalysen beschäftigt. Es folgten in einem nächsten Schritt jedoch schnell Untersuchungen zu (ii) und (iii). Neueste, in der Entwicklung noch lange nicht abgeschlossene Methoden umfassen die Punkte (iv) und (v). Insbesondere (v) bedeutet eine Abspaltung vom Gebiet der Mikroökonometrie und firmiert heute unter dem Begriff "Finanzmarktökonometrie".

Statt von den Methoden oder der Art der Daten an die Systematisierung der Mikroökonometrie zu gehen und nach neueren Entwicklungen zu fragen, können auch inhaltlich-ökonomische Aspekte den Ausgangspunkt bilden. Alle Gebiete der angewandten Mikroökonomie sind prinzipiell geeignet, mit mikroökonometrischen Methoden bearbeitet zu werden. Trotz allem zeigen sich ganz wesentliche Unterschiede. Einzug hat die Mikroökonometrie zunächst bei der Güternachfrage gehalten. Der Nachfrage nach Wohnungen und Wohneigentum wird auch heute noch besondere Aufmerksamkeit geschenkt. Die Arbeitsökonomik nimmt eine herausgehobene Position ein. Frühzeitig hat sie sich bei der empirischen Analyse mikroökonometrischer Methoden bedient und diese ganz wesentlich weiterentwickelt. Die Anfänge sind bei Untersuchungen zum Arbeitsangebot zu finden. In der Zwischenzeit werden nahzu alle Fragestellungen aus der Arbeitsökonomik mikroökonometrisch angegangen, egal ob es um Entlohnungsformen, Arbeitsplatzwechsel, Arbeitslosigkeit, Weiterbildung, Partizipation, Überstunden oder Fehlzeiten geht.

Ein starkes Standbein hat die Mikroökonometrie jedoch in der Gesundheits- und Industrieökonomik. Finanzmarktanalysen wurden bereits genannt. Hier ist eine eigenständige Entwicklung in den letzten Jahren festzustellen. In gewisser Weise gilt dies auch für die Produktions- und Kostentheorie. Unter dem Stichwort "Frontier Production Analysis" ist eine spezielle Betrachtungsweise in der Empirie zu erkennen, die zwar übliche mikroökonometrische Schätzmethoden verwendet, aber bei der Modellierung eigene Wege geht. Die Bewertung der Wirtschaftspolitik war ganz im Zuge der allgemeinen Entwicklung der Ökonometrie zunächst rein makroökonomisch ausgerichtet und hat sich in den letzten 15-20 Jahren mit Verve mikroökonometrischen Evaluationsmethoden zugewandt und auf diesem Gebiet die Methodik ganz erheblich vorangebracht. Noch nicht so deutlich ist die Handschrift der Mikroökonometrie bei der Empirie des Unternehmenswachstums sowie des Export- und Umweltverhaltens zu erkennen.

Neuere Entwicklungen in der Mikroökonometrie lassen sich auf allen drei Ebenen ausmachen, d.h. bei den Methoden und Daten sowie bei der Anwendung in den verschiedenen Teilgebieten der Ökonomie. Vorrangig geht es im Folgenden um die Methoden, die jedoch zweifelsohne zum großen Teil aufgrund inhaltlicher Fragestellungen und der Verfügbarkeit von Daten entstanden sind. Spezifikation und Schätzmethoden stehen im Vordergrund. Dem Testen wird nur eine nachrangige Aufmerksamkeit geschenkt.

Fortschritte in der Mikroökonometrie sind in allen Bereichen erkennbar, sei es in in der allgemeinen Schätz- und Testtheorie, der Modellierung, der Handhabung spezieller ökonometrischer Probleme oder sei es in der angewandten Mikroökonometrie, bei der auf verschmutzte Daten Rücksicht zu nehmen ist und inhaltliche Aspekte zu beachten sind. Der traditionelle Bereich der Mikroökonometrie kann als weitgehend abgeschlossen angesehen werden, soweit nicht auf spezielle Schätzer oder Datenarten abgestellt wird. Erwähnenswert ist die auf Vijverberg (2000) zurückgehende Verallgemeinerung, die Probit- und Logit-Modelle einschließt. Ebenso liefert die Darstellung über die Zusammenhänge zwischen bivariaten und multinomialen Modellen (Weeks/Orme 1998). Vernachlässigt wurden bisher auch Effekte, die in multinomialen Logitmodellen von alternativenspezifischen Regressoren ausgehen, die kaum variieren. Diese Lücke schließt Ronning (2002).

Als Zeitraum für die im folgenden betrachteten Entwicklungen werden ungefähr die letzten 10 Jahre zugrunde gelegt. Für einen Überblick, der die Phase davor im Blickfeld hat, sei auf König/Lechner (1994) verwiesen. Eine umfassende Darstellung sehr wichtiger Beiträge findet sich im letzten Handbook of Econometrics (Heckman/Leamer 2001). Aus angewandter Sicht für den Bereich Arbeitsökonomik zeigen Angrist und Krueger (1999) sowie Moffitt (1999) auf, wo mikroökonometrisch die Innovationen der letzten Jahre liegen. Die Einschätzungen darüber, welchen der neuen Ansätze besondere Bedeutung zukommt, sind keineswegs einheitlich. Erst in einigen Jahren lässt sich dies beurteilen, wenn die Spreu vom Weizen getrennt ist, wenn sich einige Methoden im Anwendungsbereich allgemein durchgesetzt haben, während andere als nicht praktikabel wieder in der Versenkung verschwunden sind. Hausman (2001) hat in einem kurzen Essay vor allem folgende Felder neuerer mikroökonometrischer Methoden benannt, die für die angewandte Forschung bedeutsame Implikationen besitzen: semi- und nichtparametrische Modelle der Konsumgüternachfrage, Paneldaten aus dem Bereich der Supermarktscannerdaten, Instrumentalvariablenansätze mit schwachen Instrumenten, Treatmenteffekte von Politikinterventionen, Fehler in den Variablen, vor allem nichtlineare Fehler, Quantilsregressionen und dynamische Optimierung, angewandt auf Mikrodaten. Diese Einschätzung ist gleichermaßen subjektiv wie erkenntnisreich.

Im Folgenden wird ein allgemeiner Überblick über neuere Entwicklungen in vier Bereichen gegeben: Schätzmethoden Instrumentalvariablenansätze, Paneldatenmodelle sowie semi- und nichtparametrische Modelle. Die Auswahl der vier genannten Gebiete ist durchaus subjektiv geprägt. Insgesamt bleibt aber festzuhalten, dass sich praktisch und theoretisch arbeitende Ökonometriker in die gleiche Richtung bewegen, weg von Ansätzen, die vielen Restriktionen unterliegen, hin zu stärker robusten, aber gleichzeitig einfachen, praktisch leicht handhabbaren Methoden. Unberücksichtigt in diesem Überblick bleiben Entwicklungen, die sich auf Evaluationsansätze, Finanzmarktökonometrie, Unit-Roots bei Paneldaten oder Quantilsschätzer beziehen.

2 Neuere Schätzmethoden

Ausgehend von den drei klassischen Schätzmethoden - Methode der kleinsten Quadrate, Methode der Momente und Maximum-Likelihood-Methode (KQ, MM, ML) - ist eine Reihe von Varianten und Weiterentwicklungen zu erkennen, die nur zum Teil spezifisch für die Mikroökonometrie sind. Ausgangspunkt war in vielen Fällen entweder die Überlegung, dass für Anwendungen die Bedingungen der klassischen Schätzansätze nicht erfüllt sind, sich keine analytischen Lösungen finden lassen oder eine verallgemeinernde Schreibweise von KQ-, MM- oder ML-Schätzern. Ein Beispiel für letzteren Fall sind GMM-Schätzer (generalized method of moments). Für den ersten Fall sind **Quasi-ML-Schätzer** (QML) zu nennen, auch als Pseudo-ML-Schätzer bezeichnet (Gourieroux/Monfort 1993). Wenn die (bedingte) Verteilung insgesamt nicht bekannt ist oder sich numerisch nicht handhaben lässt, aber Annahmen über die (bedingten) ersten und zweiten Momente, d.h.

$$E_0(y|x) = m(\theta_0, x)$$

und

$$V_0(y|x) = \tilde{m}(\theta_0, x),$$

sinnvoll sind, lassen sich in diesem semiparametrischen Rahmen zwar keine Likelihood-Funktionen, wohl aber Pseudo-Likelihood-Funktionen definieren. Letztere lauten

$$l_P = \Pi_{i=1}^N f(y_i; m(\theta, x_i); \tilde{m}(\theta, x_i)).$$

Sie basieren auf einer allgemeinen Klasse von Dichtefunktionen, die nicht notwendigerweise der wahren Dichte entsprechen, die aber kompatibel mit den Annahmen über die Momente sind. Unterstellt wird, dass der bedingte Erwartungswert und/oder die bedingte Varianz aus einer allgemeinen parametrischen Familie entstammt. Aus diesen Annahmen folgen konsistente und asymptotisch normalverteilte Schätzer selbst bei fehlspezifizierten Modellen. Lediglich die asymptotische Effizienz geht gegenüber üblichen ML-Schätzern verloren.

Sowohl der KQ- und als auch der ML-Schätzer basiert auf einem Optimierungskalkül. Daraus lässt sich die verallgemeinerte Form eines **Extremum-Schätzer**s herleiten

$$\hat{\theta} = argmax[m(\theta, y, x)],$$

wobei θ dem Parametervektor und $m(\theta, y, x)$ der Schätz(ziel)funktion entspricht. Eingeschlossen sind damit sowohl lineare als auch nichtlineare Modelle. Im linearen Fall führt dies Vorgehen für die Bedingungen erster Ordnung zu

$$\sum_{i=1}^{N} x_i(y_i - x_i'\beta) = 0$$

und entspricht dem OLS-Schätzer. Spezialfälle der allgemeinen Schätzfunktion führen u.a. zu robusten M-Schätzern (Judge u.a 1985, 830f, Mittelhammer u.a. 2000, 135). Ebenso kann ein Extremum-QML-Schätzer (QML-E) definiert werden (Mittelhammer u.a. 2000, 248ff). Wenn statt über die "first order conditions" direkt die Schätzfunktion $m(\theta, y, x)$ gleich Null gesetzt wird

$$m(\theta, y, x) = 0,$$

kann der QML-E-Schätzer implementiert werden. Die Schätzfunktion $m(\theta, y, x)$ bildet den Ausgangspunkt für den **EE-Ansatz** (estimating equation). Aus dieser Schätzgleichung lassen sich die Stichprobeneigenschaften ableiten. Es bedarf einer guter, möglichst optimalen Spezifikation von $m(\theta, y, x)$. Bei linearen Schätzfunktionen wird

$$m(\theta, y, x) = \eta(\theta, x)[y - g(\theta, x)]$$

gebildet (Mittelhammer u.a. 2000, 259), wobei $\eta(\theta, x)$ als unabhängig von y definiert ist. Im Falle linearer Regression folgt

$$m(\theta, y, x) = -2N^{-1}X'[y - X\beta]$$

und bei nichtlinearer Regression

$$N^{1/2} \cdot m(\theta, y, x) = 2N^{-1/2}\frac{\partial g(x, \beta)}{\partial \beta}[y - g(\theta, x)].$$

Eine Verallgemeinerung analog dem Übergang von der OLS- zur GLS-Schätzung führt zum **GEE-Schätzer** (generalized estimating equations). Dabei handelt es sich um ein multivariates Analogon der Quasi-Likelihood, oder anders ausgedrückt, um einen um die Inverse der Kovarianzmatrix erweiterten EE-Ansatz über L Gleichungen (Diggle/Liang/Zeger 1994, 144)

$$S_\theta = \sum_{l=1}^{L} \frac{\partial g_l(x, \theta)'}{\partial \theta} V(y_l)^{-1}[y_l - g_l(\theta, x)] = 0.$$

Wenn $V(y_l)$ unbekannt ist, wird eine "working covariance matrix" W verwendet , die üblicherweise fehlspezifiziert ist. Angenommen wird z. B., dass Korrelationen außerhalb der Hauptdiagonalen alle gleich sind. Oder den einfachsten Fall bildet: $W^{-1} = I$. Die Gleichungen werden dann "generalized estimating equations - GEE" bezeichnet und die Lösung entspricht einer QML-Schätzung. Das Ergebnis ist nicht mehr identisch mit dem der Minimierung von S_θ, da $\partial V(y_l)/\partial\theta$ unberücksichtigt bleibt. Im Falle einer korrekten Spezifikation der Regression $g(\theta, x)$ ist der QML-Schätzer konsistent und asymptotisch normalverteilt, eine robuste Kovarianzmatrixschätzung für $\hat{\theta}$ vorausgesetzt. Für GEE-Schätzer besitzen die Eigenschaften "Konsistenz und asymptotische Normalverteilung" Gültigkeit (Shao 1999, 312ff). Sie sind eine Möglichkeit, mit dem Problem nicht handhabbarer Likelihood-Funktionen umzugehen, sei es, weil die Dichtefunktion unbekannt ist, sei es, weil ein mehrfaches Integral zu lösen ist. Diese Fälle tauchen insbesondere bei mikroökonometrischen Ansätzen auf. Zu nennen sind vor allem multinomiale Probitmodelle, aber auch Tobitmodelle.

Alternativen zu den GEE-Schätzern bilden **simulationsbasierte Schätzer**, die im Gegensatz zu den GEE-Schätzern bereits relativ häufig zum Einsatz kommen. Prinzipiell wird zwischen vier simulationsbasierten Schätzmethoden unterschieden (Stern 1997, 2029ff, Gourieroux/Monfort 1996):

- Methode der simulierten Likelihood (MSL)

$$lnL = \sum_{i=1}^{N} ln\hat{L}_i(\theta|y_i, X_i)$$

- Methode der simulierten Momente (MSM)

$$Q'\hat{h}(y, X|\theta) = 0$$

- Methode der simulierten Scores (MSS)

$$\frac{\partial lnf(y_i|X_i, \hat{\theta})}{\partial \theta} = 0$$

- Monte-Carlo-Markov-Ketten-Methode (MCMC)

$$f(dy|X, \theta) = \int P(dy, \tilde{y}|X, \hat{\theta}) \cdot f(\tilde{y}|X, \hat{\theta})d\tilde{y}.$$

Die dritte Methode ist ein Spezialfall der zweiten, indem die allgemeinen Bedingungen an Momente auf Bedingungen erster Ordnung beschränkt werden. Wie allgemein bekannt ist, kann der ML-Schätzer als GMM-Schätzer betrachtet werden. Letzterer basiert auf den Schätzbeschränkungen, verbunden mit den Scorefunktionen. Lassen sich die Likelihoodbeiträge der einzelnen Beobachtungen $L_i(\theta|y_i, X_i)$, die allgemeine Funktion für die Momentenbedingung $h(y, X|\theta) = 0$, die ersten Ableitungen der Log-Likelihoodfunktion $\frac{\partial lnf(y_i|X_i,\hat{\theta})}{\partial \theta} = 0$ oder die Übergangswahrscheinlichkeit $P(dy, \tilde{y}|X, \hat{\theta})$ von der (angenommenen) Stichprobenverteilung $f(\tilde{y}|X, \hat{\theta})$ zur wahren, aber unbekannten Verteilung $f(y|X, \theta)$ im Sinne von Markov-Ketten analytisch nicht bestimmen, besteht die Möglichkeit, auf simulierte Werte zurückzugreifen. Es existieren verschiedene Simulationsmethoden (Simulatoren), u.a.

- einfacher Häufigkeitssimulator (CFS)
- Importance-Sampling-Simulator (ISS)
- Stern-Dekompensationssimulator (SDS)
- Geweke-Hajivassiliou-Keane-Simulator (GHKS)

(Börsch-Supan 1994, Stern 1997), die auf die ersten drei Methoden angewandt werden. Dem stehen zwei üblicherweise verwendete Ansätze für die MCMC-Methode gegenüber, der Gibbs-Algorithmus und der Metropolis-Hastings-Algorithmus (Chib 2001, Geweke/Keane 2001, Gilks/Richardson/Spiegelhalter (1997), Robert/Casella 2000). Anwendungsbeispiele bieten unter anderem Börsch-Supan (1994), Czado (2000), Inkmann (2001), Kaltenborn (1997), Mühleisen (1994), Ziegler/Eymann (2001). Der Häufigkeitssimulator (crude frequency simulator - CFS) ist der einfachste diskrete Simulator. Die Wahrscheinlichkeit für ein Ereignis wird dadurch approximiert, dass ausgezählt wird, wie häufig dieses Ereignis E bei einer Häufigkeitsverteilung in einer Replikation auftritt, bezogen auf die Gesamtzahl der Elemente dieser Replikation, um dann den Durchschnitt über alle Replikationen zu bilden. Im Falle der MSL wird der CFS als arithmetisches Mittel der simulierten Häufigkeitsdichten über alle Replikationen R gebildet

$$\hat{f}(\theta) = argmax \sum_{i=1}^{N} ln[\frac{1}{R} \sum_{r=1}^{R} \hat{f}(\theta; y_i, x_i, u_i^r).$$

Dieser Simulator ist zwar erwartungstreu, aber ansonsten unbefriedigend, da er diskreten Charakter besitzt und einen sehr hohen Rechenaufwand verlangt. Das erste Problem lässt sich durch einen Kernschätzer lösen, allerdings auf Kosten der Erwartungstreue. Der ISS benutzt zwei Verteilungen f(u) und g(u) und berechnet den Mittelwert

$$I(u \in E) \cdot \frac{f(u)}{g(u)}$$

mit $I(\cdot)$ als Indikatorvariable. Die beiden restlichen Simulatoren (SDS, GHKS) sind Spezialfälle des ISS. Stern nutzt die Eigenschaft, dass die Summe von zwei normalverteilten Zufallsvariablen selbst normalverteilt ist und zerlegt die Störgröße in einen homoskedastischen Teil und einen davon unabhängigen normalverteilten Rest mit positiv definiter Kovarianzmatrix (Stern 1997, 2027). Am häufigsten Verwendung findet der GHK-Simulator, der, wie Simulationsstudien zeigen, dem SDS hinsichtlich des MSE überlegen ist (Börsch-Supan/Hajivassiliou 1993). Der GHKS ist charakterisiert durch das Produkt bedingter univariater Wahrscheinlichkeiten, d.h. anstelle des Integrals

$$P = \int_{\infty}^{b} f(u)du$$

wird b in M Unterabschnitte zerlegt

$$P = \Phi(b_1) \cdot \Phi(b_2|u_1) \cdot \Phi(b_3|u_1, u_2) ... \cdot \Phi(b_M|u_1, u_2 ... u_{M-1}).$$

Es erfolgt eine sequenzielle Ziehung aus einer gestutzten Normalverteilung, wobei in den nachfolgenden Ziehungen die Ergebnisse der vorangegangenen Ziehungen als Bedingungen eingehen. Für den **MCMC-Algorithmus** ist für

$$f(dy|X, \theta) = \int P(dy, \tilde{y}|X, \hat{\theta}) \cdot f(\tilde{y}|X, \hat{\theta}) d\tilde{y}.$$

ein geeignetes $P(dy, \tilde{y}|X, \hat{\theta})$ zu finden, wobei $f(dy|X, \theta)$ eine invariante Verteilung ist und $f(dy|X, \theta) = f(y|X, \theta)dy$ gilt. Dies mag wie die Suche nach der Nadel im Heuhaufen erscheinen. Das Verfahren vereinfacht sich jedoch, wenn angenommen wird, dass die Übergangswahrscheinlichkeit (transition kernel) ausgedrückt werden kann durch

$$P(dy, \tilde{y}|X, \hat{\theta}) = p(y, \tilde{y}|X, \hat{\theta})dy + r(\tilde{y}|X, \hat{\theta}) \delta_{\tilde{y}|X, \hat{\theta}} dy.$$

Daraus lässt sich die Reversibilitätsbedingung

$$f(\tilde{y}|X,\theta) \cdot p(\tilde{y},y|X,\theta) = f(y|X,\theta) \cdot p(y,\tilde{y}|X,\theta)$$

ableiten. Für $p(y,\tilde{y}|X,\theta)$ sind geeignete Kandidaten $q(y,\tilde{y}|X,\theta)$ zu suchen, die diese Bedingung erfüllen. Üblicherweise wird jedoch

$$f(\tilde{y}|X,\theta) \cdot q(\tilde{y},y|X,\theta) > f(y|X,\theta) \cdot q(y,\tilde{y}|X,\theta)$$

eintreten. Zur Korrektur ist nach dem **Metropolis-Hastings-Algorithmus** (Chib/Greenberg 1995, Chib 2001, 3580ff, spez. 3583) eine Wahrscheinlichkeit $\alpha(y,\tilde{y}) < 1$ einzuführen, so dass

$$f(\tilde{y}|X,\theta) \cdot q(\tilde{y},y|X,\theta) \cdot \alpha(\tilde{y},y) = f(y|X,\theta) \cdot q(y,\tilde{y}|X,\theta) \cdot \alpha(y,\tilde{y}) = f(y|X,\theta) \cdot q(y,\tilde{y}|X,\theta)$$

folgt. Der Metropolis-Hastings-Schätzer für $p_{MH}(y,\tilde{y}|X,\theta)$ lautet damit

$$p_{MH}(y,\tilde{y}|X,\theta) \equiv q(\tilde{y}|X,\theta)\alpha(d\tilde{y},y|X,\theta) = q(\tilde{y}|X,\theta) \cdot min[\frac{f(y|X,\theta) \cdot q(y,\tilde{y}|X,\theta)}{f(\tilde{y}|X,\theta) \cdot q(\tilde{y},y|X,\theta)},1].$$

Wird zusätzlich

$$r(\tilde{y}) = 1 - \int q(\tilde{y}|X,\theta)\alpha(d\tilde{y},y|X,\theta)dy$$

berücksichtigt, dann folgt damit die Übergangswahrscheinlichkeit nach Metropolis-Hastings $P(y,\tilde{y}|X,\theta)_{MH}$. Die Bestimmung von $\alpha(\tilde{y},y)$ verlangt nicht die genaue Kenntnis von $f(\cdot)$, da diese Funktion im Zähler und im Nenner von α auftaucht. Ein Element y aus $q(y|X,\theta)$ wird im Sinne der Reversibilitätsbedingung akzeptiert, wenn $\alpha(\tilde{y},y) \geq u$, wobei u ein zufällig entnommenes Element der gleichverteilten Zufallsvariablen $U \sim UNI(0,1)$ ist. Gezogen werden N Elemente $(\tilde{y}^{(1)}...\tilde{y}^{(N)})$. Falls $\alpha(\tilde{y}^{(j)},y) \leq u$, wird $\tilde{y}^{(j)}$ gleich dem nächsten (akzeptierten) Element gesetzt $(\tilde{y}^{(j)} = \tilde{y}^{(j+1)})$.

Beim **Gibbs-Algorithmus** wird der Parametervektor θ in P Blöcke zerlegt (Chib 2001, 3589ff) und jeder Block wird auf Basis der vollen bedingten Verteilung dieses Blocks gezogen, das heißt, wenn θ_p der betrachtete Block ist, dann gehen alle anderen $\theta_1...\theta_{p-1}\theta_{p+1}...\theta_P$ als Bedingung ein, notiert als $f(\theta_p|\theta_{-p})$. Der volle Satz an bedingten Verteilungen lautet damit

$$\{f(\theta_1|\theta_2,...,\theta_P); f(\theta_2|\theta_1,\theta_3,...,\theta_P); ...; f(\theta_P|\theta_1,\theta_3,...,\theta_{P-1})\}.$$

Rekursiv erfolgt dann ein Update der bedingten Dichtefunktionen. Für alle θ-Blöcke, für die der Index kleiner ist als der gerade betrachtete p, gehen die auf der gleichen Stufe bereits aktualisierten Blöcke als Bedingung ein, während für die restlichen noch auf der Ergebnisse der vorangegangenen Stufe als Bedingung zurückgegriffen werden muss. Der Gibbs-Algorithmus lässt sich damit in folgenden Schritten beschreiben:

(1) Wähle einen Vektor von Anfangswerten $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, ..., \theta_P^{(0)})$

(2) Wiederhole in M Stufen und jeweils P Schritten die Aktualisierung von θ. Auf der m-ten Stufe bedeutet dies, es sind nacheinander zu generieren

$$\begin{array}{lll} \theta_1^{(m)} & aus & f(\theta_1|\theta_2^{m-1},...,\theta_P^{m-1}) \\ \theta_2^{(m)} & aus & f(\theta_2|\theta_1^{m},\theta_3^{m-1},...,\theta_P^{m-1}) \\ \vdots & & \\ \theta_P^{(m)} & aus & f(\theta_P|\theta_1^{m},\theta_2^{m},...,\theta_{P-1}^{m}). \end{array}$$

Die Übergangsdichte der Markowkette eines Parameterblocks der Stufe m zu m+1 wird bestimmt durch

$$f(\theta^m, \theta^{m+1}) = \prod_{p=1}^{P} f(\theta_p | \theta_1^{(m+1)}, ..., \theta_{p-1}^{(m+1)}, \theta_{p+1}^{(m)}, ..., \theta_P^{(m)}).$$

3 Instrumentalvariablenansätze

In der Ökonometrie wurde bereits früh erkannt, dass sich das Problem inkonsistenter Schätzungen bei Abhängigkeit von Regressoren und Störgrößen durch Instrumentalvariablenschätzungen (IV-Schätzungen) lösen lässt (Wright 1928). Abhängigkeiten resultieren vor allem aus simultanen Zusammenhängen, Zufallsfehlern bei der Messung exogener Variablen und aus der Berücksichtigung von verzögerten endogenen Variablen als Bestimmungsfaktoren bei gleichzeitigem Auftreten von Autokorrelation. Die IV-Methode ist stets Gegenstand ökonometrischer Auseinandersetzungen geblieben (Bowden/Turkington 1984) und erlebt gegenwärtig eine Renaissance, jedoch mit einer etwas anderen Akzentuierung als bisher. In der letzten Zeit werden Instrumentalvariablen im Zusammenhang mit Evaluationsmethoden, mit der Erfassung von Treatmenteffekten bei unterdrückten, nicht beobachteten Einflussgrößen erörtert. Es ist schwierig alle wesentlichen Einflussgrößen zu erfassen. Dies allein stellt noch kein Problem dar. Wenn aber die nicht erfassten Determinanten sowohl auf die endogene Variable als auch auf Regressoren Einfluss nehmen, so werden mit üblichen Verfahren die Effekte der erfassten exogenen Variablen verzerrt und inkonsistent geschätzt. Wenn es um Treatmenteffekte geht, besteht eine Lösung des Problems darin, eine Zufallsauswahl von Merkmalsträgern für die Teilnehmer- und Kontrollgruppe vorzunehmen. Dann sollten unbeobachtete Variablen keine verzerrenden Effekte hervorrufen. Abgesehen davon, dass diese Ausssage keine Allgemeingültigkeit besitzt (Hübler 2001a, S. 116f), sind Zufallsexperimente nicht immer möglich. Instrumentalvariablen bieten auch hier eine mögliche Lösung durch sogenannte "natürliche Experimente". Gemessen wird der Behandlungseffekt durch den von Imbens und Angrist (1994) als "Local Average Treatment Effect (LATE)" bezeichneten Indikator

$$\frac{E(y_i|Z=1) - E(y_i|Z=0)}{E(X|Z=1) - E(X|Z=0)},$$

wenn die erklärende Variable X und die Instrumentalvariablen Z Dummy-Variable sind. Auf Einzelheiten hierzu wird in diesem Beitrag nicht eingegangen. Besonders durch diese neuere Diskussion über Instrumentalvariablen hat die Methode Bedeutung für die Mikroökonometrie gewonnen.

3.1 IV-, MM- und GMM-Schätzer

Allgemeines Grundanliegen der **IV-Methode** ist, für einzelne Regressoren aus einer Regressionsgleichung $y = X\beta + u$, z.B. für x_K, Ersatzvariablen $z_{K1}, \cdots, z_{Kl}$ (Instrumentalvariablen) zu finden, die hoch korreliert sind mit x_K, aber unkorreliert mit der Störgröße u. Neben der Designmatrix $X \sim N \times K$ findet die Instrumentalvariablenmatrix $Z = (x_1, \cdots, x_{K-1}, z_{K1}, \cdots, z_{Kl}) \sim N \times (N \cdot (K-1+l)$ Eingang in Schätzungen. Die zentrale Bedingung der Unkorreliertheit

(1) $$E(Z'u) = 0,$$

die sich nicht beobachten lässt, wird durch die analoge Stichprobenbedingung

$$\frac{1}{N}Z'\hat{u} = \frac{1}{N}Z'(y - X\hat{\beta}) = 0 \tag{2}$$

ersetzt, wie dies bei der **Methode der Momente** (MM) geschieht. Momente der Grundgesamtheit werden durch Momente der Stichprobe geschätzt. Die notwendige Schätzung für β lässt sich nur dann durch einfaches Auflösen über

$$\hat{\beta} = (Z'X)^{-1}Z'y \tag{3}$$

erreichen, wenn $Z = (x_1, \cdots, x_{K-1}, z_K)$. Andernfalls kann die Ausgangsgleichung transformiert

$$Z'y = Z'X\beta + Z'u \tag{4}$$

und statt $u'u$ der Ausdruck $u'ZZ'u = (y-X\beta)'ZZ'(y-X\beta)$ minimiert werden. Dieses Vorgehen führt zu

$$\hat{\beta} = (X'ZZ'X)^{-1}X'ZZ'y. \tag{5}$$

Der Schätzer ist konsistent, aber nicht effizient. Eine allgemeinere Form der Schätzung ergibt sich, wenn eine gewichtete Distanz minimiert wird

$$(y - X\beta)'ZWZ'(y - X\beta) = Min., \tag{6}$$

wobei W eine Gewichtungsmatrix ist. Der daraus resultierende Schätzer wird **verallgemeinerter Momentenschätzer** (GMM) genannt, da die verallgemeinerte Momentenbedingung

$$E[(Z'X)'WZ'u] = 0$$

durch die analoge Stichprobenmomentenbedingung ersetzt wird. Hansen (1982) zeigt, dass $\hat{W} = \hat{V}^{-1}$ optimal ist, wenn $\hat{W}$ einer konsistenten Schätzung für $V((1/N)Z'u)^{-1}$ entspricht. Die Stichprobenmomentenbedingung lautet dann

$$\frac{1}{N}X'Z\hat{W}Z'\hat{u} = \frac{1}{N}(X'Z)\hat{V}^{-1}(Z'y - Z'X\hat{\beta}) = 0.$$

Als Schätzung für β folgt daraus

$$\hat{\beta} = (X'Z\hat{V}^{-1}Z'X)^{-1}(X'Z\hat{V}^{-1}Z'y). \tag{7}$$

Wenn $\hat{W} = I$, liegt der Schätzer aus (5) vor. Im Fall von Homoskedastie und Abwesenheit von Autokorrelation ist $\hat{V} = \frac{1}{N}\hat{\sigma}^2(Z'Z)^{-1}$. Dies entspricht dem 2SLS-Schätzer

$$\hat{\beta}_{2SLS} = (X'Z(Z'Z)^{-1}Z'X)^{-1}(X'Z(Z'Z)^{-1}Z'y). \tag{8}$$

GMM-Schätzer sind also eine Verallgemeinerung der IV-Schätzer, die als Momentenschätzer zu interpretieren sind und unter denen der 2SLS-Schätzer ein Spezialfall ist. Weitere bekannte Spezialfälle sind der LIML-, k-Klassen-, 3SLS- und Jackknife-IV-Schätzer. Sowohl IV- als auch GMM-Schätzer besitzen günstige asymptotische Schätzeigenschaften. GMM-Schätzer sind konsistent und asymptotisch normalverteilt. Die Effizienz hängt aber wesentlich von der Wahl der Kovarianzmatrix ab. Bei kleinen Stichproben können ganz erhebliche Verzerrungen der GMM-Schätzer resultieren (Altonji/Segal 1996). Andrews (1999) trägt mit seinem Vorschlag zur Auswahl konsistenter Momente zu einer Verbesserung der GMM-Schätzer bei. Daneben sind informationstheoretische Alternativen zu den GMM-Schätzern eingeführt worden (Imbens/Spady/Johnson 1998, Nevo 2002).

3.2 Kleine-Stichprobeneigenschaften und Kritik an IV-Schätzern

Über kleine Stichprobeneigenschaften ist weniger bekannt. Sie unterscheiden sich zwischen den einzelnen IV-Schätzern. Nelson und Startz (1990) haben die Verteilung von IV-Schätzern bei kleinen Stichproben untersucht und sind zu folgenden Eigenschaften gekommen: IV-Schätzer sind verzerrt. Die Verteilung ist bimodal mit relativ starken Besetzungen in den Randbereichen und die Schätzwerte können sich um einen Punkt konzentrieren, der häufig näher beim OLS-Schätzer als beim wahren Parameter liegt. Auch Bound, Jaeger und Baker (1995) haben darauf aufmerksam gemacht, dass bei endlichen Stichproben die Verzerrung in die gleiche Richtung wie beim OLS-Schätzer geht. Und zwar ist die Annäherung um so größer, je mehr das partielle Bestimmtheitmaß zwischen z und x gegen Null tendiert. Selbst bei großen Stichprobenumfängen ist die Gefahr verzerrter IV-Schätzer nicht auszuschließen. Im Extremfall kann sie sogar größer als bei der OLS-Schätzung sein. Wenn Regressoren und Störgröße unkorreliert sind, dann entspricht die Verteilung der IV-Schätzer der exakten Verteilung.

Bestimmtheitsmaße in IV-Modellen bedürfen einer gesonderten Erörterung. Sie unterscheiden sich von denen in üblichen Regressionsmodellen, auch wenn Programmpakete zum Teil routinemäßig die gleichen ausweisen. Pesaran und Smith (1994) diskutieren, warum die üblichen R^2-Werte ungeeignet sind und entwickeln einen neuen verallgemeinerten Goodness-of-Fit-Indikator. Dieser Indikator basiert auf dem zweistufigen Ansatz der IV-Schätzung und verwendet das Bestimmtheitsmaß der zweiten Stufe. Es wird also z.B. nicht direkt über (7) geschätzt, sondern in der ersten Stufe wird die Kovarianzmatrix ermittelt, um dann in der zweiten Stufe die IV-Matrix Z durch $\hat{Z} = C_W Z = W(W'W)^{-1}W'Z$ mit $W = Z\hat{V}^{-1}Z'$ zu ersetzen. Die daraus resultierenden Residuen weichen von denen aufgrund von (7) ab. Letztere wäre vorzuziehen. In diesem Fall lässt sich jedoch keine Zerlegung der Streuung von y in die erklärte und nicht erklärte Varianz vornehmen. Windmeijer (1995) vergleicht dieses Maß mit einem zweiten, das versucht, R^2 direkt zu bestimmen. Beim direkten IV-Schätzer steht man vor dem Problem, dass das Ausgangsmodell kein absolutes Glied besitzt. Es wird deutlich, wie dies auch aus der Diskussion um Goodness-of-Fit-Indikatoren bei nichtskalaren Kovarianzmatrizen bekannt ist (Judge/Griffiths/Hill/Lütkepohl/Lee 1985, S.31ff.), dass kein allgemein bestes Goodness-of-Fit-Maß existiert.

Kritik an IV-Schätzern entzündet sich an verschiedenen Punkten. Zentral ist die Wahl der Instrumente. Wie sind Variablen zu ermitteln, die einerseits mit dem Regressoren stark und mit der Störgröße nicht korreliert sind? Während früher vor allem verzögerte exogene und endogene Variablen als Instrumente verwendet wurden, konzentriert sich heute die Wahl mehr auf aktuelle und weiter zurückliegende Differenzen exogener Variablen, auf Gruppenmittelwerte und insbesondere auf Wissen über ökonomische Mechanismen und institutionelle Gegebenheiten. Bekannte Beispiele für letzteren Fall sind die Losnummer, die Wehrpflichtige in den USA während des Vietnamkrieges gezogen haben und nach denen zur Armee eingezogen wurde, als Instrument für den Veteranenstatus, das Quartal, in dem Personen geboren wurden, als Instrument für die Zahl der Schuljahre, angekündigte Entlassungen und berufsspezifische Arbeitsmarktchancen als Instrumente für Mobilität. Die über die Maimonides-Regel geschätzte Klassenstärke als Instrumentalvariable für die tatsächliche Zahl an Schüler als Bestimmungsgrund für das Testergebnis von Schülern ist ein weiteres Beispiel, das Angrist und Lavy (1998) vorstellen. In öffentlichen Schulen in Israel werden Klassen nach einem Vorschlag von Maimonides geteilt, wenn sie mehr als 40 Schüler haben. Beträgt in einer Schule s die Zahl der

Schüleranmeldungen b_s, so wird als Prognose für die tatsächliche Klassenstärke der i-ten Klasse in der s-ten Schule

$$z_s = \frac{b_s}{int(\frac{b_s - 1}{40}) + 1}$$

als Instrumentalvariable gewählt mit int als Integerzahl. Eine übliche Kritik an dem Vorgehen, bei dem die Instrumente aufgrund einer institutionellen Regelung bestimmt werden, ist, dass die Wahl der Instrumente kaum direkt theoriegeleitet erfolgt (Rosenzweig/Wolpin 2000). Zwar werden die Instrumente üblicherweise nicht aus einem formalen Modell abgeleitet, der theoretische und institutionelle Hintergrund, der die Wahl der Instrumente bestimmt, ist jedoch häufig weit stärker ausgeprägt als bei einem direkten Bezug auf einen einfachen Modellansatz.

Schwerwiegender ist der Einwand, dass die Instrumente z möglicherweise nur bei einem Teil der untersuchten Personen Einfluss auf x nehmen, wenn heterogene Reaktionen von z auf x vorliegen, wenn sich z.B. ältere Beschäftigte mit langer Betriebszugehörigkeitsdauer von einer Entlassungsankündigung nicht beeindrucken lassen. Dann ist für diese Gruppe die Entlassungsankündigung kein geeignetes Instrument für individuelle Mobilität. In solchen Fällen ist es sinnvoll, verschiedene "natürliche Experimente" durchzuführen. Also in dem genannten Beispiel nicht nur die Entlassungsankündigung z_1, sondern auch die beruflichen Arbeitsmarktchancen z_2 als IV heranzuziehen. Grundgedanke dieser Überlegung ist, dass z_1 und z_2 heterogene Reaktionen nicht bei den gleichen Gruppen hervorzurufen.

3.3 Schwache Instrumente bei IV-Schätzern

Schlechte Instrumente liegen immer dann vor, wenn Z und u korrelieren, wenn also auch die Instrumente nicht unabhängig von unterdrückten Variablen sind. Genauso negative Folgen haben jedoch "schwache" Instrumente, das heißt, wenn z nur einen geringen Teil der Variation von x erklären kann. Bound, Jaeger und Baker (1995) haben aufgezeigt, dass in diesen Fällen bei IV-Schätzern eine Tendenz zur Inkonsistenz besteht, selbst wenn Z und u unkorreliert sind. Die asymptotische Verteilung des IV-Schätzers ist bei Abhängigkeit von X und u nur eine schlechte Approximation der wahren Verteilung, wenn es sich um schwache Instrumente handelt. Han und Schmidt (2001) weisen nach, dass der Erwartungswert des IV-Schätzers dem Wahrscheinlichkeitslimes des OLS-Schätzers entspricht, wenn Regressoren und Instrumente unkorreliert sind. Der IV-Schätzer ist dann weder konsistent noch asymptotisch normalverteilt. Eine verbesserte Inferenztechnik bei schwachen Instrumente gegenüber der sonst üblichen Vorgehensweise entwickeln Wang und Zivot (1998). Asymptotisch valides Testen oder die Bildung von Konfidenzintervallen für strukturelle Parameter ist mit Hilfe von LR- und LM-, statt mit Wald-Statistiken möglich und praktikabel. Diese sind χ^2-verteilt mit der Anzahl der Instrumente als Freiheitsgrade. Die Statistiken basieren auf ML- und GMM-Schätzern.

Um festzustellen, ob schwache Instrumente vorliegen, empfehlen die Autoren bei empirischen Untersuchungen, die Instrumentalvariablen verwenden, routinemäßig folgende Indikatoren auszuweisen:

- das partielle Bestimmtheitsmaß

$$R^2_{x_K;z_{K1},\cdots,z_{Kl}|x_1,x_2,\cdots,x_{K-1}}$$

zwischen Regressor x_K und Instrumentalvariablen $z_{K1},\cdots,z_{Kl}$;

- die F-Teststatistik für die Regression $X^* = Z^*\Pi + V$, wobei aus X^* und Z^* der Einfluss der in X und Z gemeinsam vorhandenen Variablen auspartialisiert ist.

Niedrige partielle R^2-Werte und niedrige F-Werte deuten auf schwache Instrumente hin. Wenn

$$F < F_l^1(1-\alpha),$$

dann ist, falls nur ein Regressor x_K betrachtet wird, dessen Instrumentierung durch die Instrumente $z_{K1}, z_{K2}, \cdots, z_{Kl}$ als ungeeignet abzulehnen. Staiger und Stock (1997) geben als Daumenregel einen Wert von 10 an. Ist $F > 10$, dann wird die Instrumentierung akzeptiert. Sie demonstrieren, dass der 2SLS-Schätzer selbst bei sehr großen Datensätzen diesem Kriterium häufig nicht genügt. Er weist einen erheblichen Bias auf, während der LIML-Schätzer, bezogen auf den Median, unverzerrt ist. Der Bias des 2SLS-Schätzers nimmt mit dem Grad der Überidentifikation zu, das heißt, je mehr Instrumente im Vergleich zu der Zahl der endogenen Regressoren vorliegen. Angrist, Imbens und Krueger (1999) schlagen als Alternative einen Jackknife-IV-Schätzer (JIV) vor, der asymptotisch mit dem 2SLS-Schätzer übereinstimmt, aber aufgrund von kleinen Stichprobeneigenschaften überlegen ist und in dieser Hinsicht dem LIML-Schätzer sehr nahe kommt.

Beim JIV-Schätzer wird auf der ersten Stufe statt $\hat{X} = Z\hat{\Pi}$ die Matrix $\hat{X}_{JIV}$ zeilenweise bestimmt mit

$$Z_i\tilde{\pi} = Z_i(Z(i)'Z(i))^{-1}(Z(i)X(i)),$$

wobei $Z(i)$ und X(i) die Z- bzw. X-Matrix ohne die Zeile für die i-te Beobachtung ist. Aufgrund dieses Vorgehens soll die Eigenschaft der Unkorreliertheit zwischen Z und u besser gesichert werden, weil die Störgröße der i-ten Beobachtung (u_i) unabhängig von X_j für $i \neq j$ ist d.h. $E(u_i Z_i\tilde{\pi}) = 0$. Oder anders ausgedrückt, der Bias des 2SLS-Schätzers, der durch die i-te Beobachtung hervorgerufen wird, soll durch die Konstruktion eines optimalen Instruments für die i-te Beobachtung reduziert werden. Die Eigenschaften des JIV-Schätzers bei endlichen Stichproben sind in Bezug auf Bias und Konvergenz denen des 2SLS-Schätzers überlegen und ähnlich denen des LIML-Schätzers.

Da die IV-Schätzer stark sensitiv auf die Zahl und Form valider Instrumente reagieren, ist es naheliegend, die Zahl der Instrumente nach einem festen Kriterium auszuwählen. Während Andrews (1999) nach der maximalen Zahl valider Instrumente sucht, bestimmen Donald und Newey (2001) aufgrund der Minimierung des MSE-Kriteriums aus einer vorgegebenen Menge valider Instrumente eine Untermenge. Durch die Optimierung der Zahl der Instrumente können die Standardfehler gegenüber dem traditionellen LIML-Ansatz um 50 Prozent reduziert werden. Hier erscheint es optimal, den größtmöglichen Satz an validen Instrumenten auszuwählen. Im Vergleich zum üblichen 2SLS-Schätzer beträgt die Reduktion des Standardfehlers sogar bis 100 Prozent. Die minimale Zahl an Instrumenten ist hier von Vorteil. Der LIML-Schätzer erweist sich dem 2SLS-Schätzer aufgrund von Simulationen nicht, wie andere Studien nahelegen, durchgängig als überlegen. Diese Dominanz ist stärker bei großen als bei kleinen Stichproben ausgeprägt. Neben LIML- und 2SLS-Schätzer untersuchen Donald und Newey auch einen biaskorrigierten 2SLS-Schätzer (B2SLS), der die kleinste Zahl an Instrumenten verwendet und dem LIML-Schätzer sehr ähnlich ist. Bei optimaler Wahl der Instrumentenzahl kann der Standardfehler ähnlich stark wie beim 2SLS-Schätzer reduziert werden.

Das sehr einfache Vorgehen zur Aufdeckung schwacher Instrumente über partielle Bestimmtheitsmaße und F-Werte ist nicht in jedem Fall befriedigend. Hahn und Hausman (2002) entwickeln ein zweistufiges Testverfahren, das sensitiver auf schwache Instrumente reagiert. Bei beiden Teiltests wird der Hausman-Testidee des Vergleichs von zwei Schätzern gefolgt. Ausgangspunkt bildet ein einfaches Zweivariablenmodell

$$y_1 = \beta y_2 + u_1 = \beta Z\pi_2 + \epsilon_1 \tag{9}$$

$$y_2 = Z\pi_2 + \epsilon_2, \tag{10}$$

wobei

$$\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix} \sim N(0,\Omega) = N\left(0, \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{bmatrix}\right).$$

Die IV-Matrix bestehe aus l Instrumenten. Geschätzt wird zweimal nach dem 2SLS-Verfahren, und zwar einerseits das Ausgangsgleichungssystem (9) und (10) und andererseits an Stelle von (9) die Umkehrregression

$$y_2 = \frac{1}{\beta} y_1 - \frac{1}{\beta} u_1 =: \gamma y_1 + \tilde{u}_1 \tag{11}$$

in Verbindung mit (10). Bei den Prognosen für y_1 und y_2 aus der Ausgangs- bzw. Umkehrregressionschätzer handelt es sich um orthogonale Projektionen in dem Unterraum, der durch Z aufgespannt wird. Daraus ergibt sich, dass unter üblicher Asymptotik erster Ordnung $\hat{\beta}_{2SLS}$ und $\hat{\gamma}_{2SLS}$ vollständig miteinander korreliert sind ($\sqrt{N}(\hat{\beta}_{2SLS} - \hat{\gamma}_{2SLS}) = o_p(1)$), auch wenn die beiden Schätzungen numerisch erheblich voneinander abweichen können (Hübler 1990a). Um einen Hausman-Test bilden zu können, wird die Varianz für die Differenz der beiden Schätzer benötigt. Die asymptotische Varianz erster Ordnung ist hierfür wegen der vollständigen Korrelation der beiden Schätzer nicht geeignet. Statt dessen wird Asymptotik zweiter Ordnung betrachtet, d.h. es wird das asymptotische Verhalten für das Verhältnis der Zahl der Instrumente zur Zahl der Beobachtungen gegen einen konstanten Wert untersucht. Bei Verwendung der asymptotischen Varianz zweiter Ordnung lässt sich, wie die Autoren zeigen, ein Grenzwert B für die Differenz der beiden Schätzer ermitteln, der in Abhängigkeit von y_1, y_2, Z, N und $\frac{l}{N} =: \alpha$ (Bekker 1994, S. 658) mit l als Zahl der Instrumente geschätzt werden kann. Als asymptotisch t-verteilte Teststatistik bei Zugrundelegung der Asymptotik zweiter Ordnung folgt

$$T_1 = \frac{\sqrt{N}(\hat{\beta}_{2SLS} - \frac{1}{\hat{\gamma}_{2SLS}} - \hat{B})}{\sqrt{\hat{V}(\hat{\beta}_{2SLS} - \frac{1}{\hat{\gamma}_{2SLS}} - \hat{B})}} =: \frac{\hat{d}_1}{\sqrt{\hat{w}_1}}, \tag{12}$$

wobei w_1 durch y_1, y_2, Z, l, N und den LIML-Schätzer für β bestimmt ist. Wird die Hypothese $H_0 : plim\sqrt{N}(\beta_{2SLS} - \frac{1}{\gamma_{2SLS}} - B) = 0$ abgelehnt, d.h. ist $T_1 < z_{\frac{\alpha}{2}}$ oder $T_1 > z_{1-\frac{\alpha}{2}}$ mit z als Prozentpunkt der Standardnormalverteilung, so können nach Hahn und Hausman zwei Gründe dafür verantwortlich sein. Entweder die Orthonalitätsannahme der Instrumente ist falsch oder die Stichprobeneigenschaften der asymptotischen Approximation erster Ordnung sind nicht hinreichend genau. Letzteres wird als Indiz für schwache Instrumente gewertet. Wird H_0 nicht abgelehnt, so liegt kein Problem mit schwachen Instrumenten vor und der 2SLS-Schätzer kann verwendet werden. Bei Ablehnung von H_0 ist der LIML-Schätzer in Erwägung zu ziehen, der, wie andere Studien gezeigt haben, bei schwachen Instrumenten besser als der 2SLS-Schätzer abschneidet. Ob der LIML-Schätzer jedoch wirklich geeignet ist, überprüfen die Autoren durch einen zweiten Test. Dieser basiert auf dem Vergleich eines biaskorrigierten Ausgangs- und Umkehrregressionsschätzer. Hierzu schlagen Hahn und Hausman die von Donald und Newey (2001) diskutierte biaskorrigierte Version des 2SLS-Schätzers (B2SLS) vor. Es folgt wiederum eine auf Asymptotik zweiter Ordnung basierende Teststatistik

$$T_2 = \frac{\sqrt{N}(\hat{\beta}_{B2SLS} - \frac{1}{\hat{\gamma}_{B2SLS}})}{\sqrt{\hat{V}(\hat{\beta}_{B2SLS} - \frac{1}{\hat{\gamma}_{B2SLS}})}} =: \frac{\hat{d}_2}{\sqrt{\hat{w}_2}}. \qquad (13)$$

Die Koeffizientenschätzer und der Varianzschätzer lassen sich aus folgenden Elementen bestimmen: $y_1, y_2, Z, l, N, \hat{\beta}_{LIML}$. Wurde die erste Hypothese abgelehnt und ist $z_{\frac{\alpha}{2}} < T_2 < z_{1-\frac{\alpha}{2}}$, dann empfehlen Hahn und Hausman die Verwendung des LIML-Schätzers, der unter den drei Schätzern 2SLS, B2SLS und LIML bei großer Anzahl an Instrumenten am besten abschneidet (Donald/Newey 2001,1165). Der LIML-Schätzer sollte allerdings weder vom B2SLS-Ausgangs- noch vom Kehrwert des B2SLS-Umkehrregressionsschätzers stark abweichen. Wenn auch die Nullhypothese zu T_2 abgelehnt wird, muss auf extrem schwache Instrumente geschlossen werden, so dass dann auch der LIML-Schätzer nicht zu empfehlen ist. Eine Alternative, mit dem Problem schwacher Instrumente umzugehen, könnte dann in dem Vorschlag von Angrist und Krueger (1995) bestehen, mit IV-Schätzungen für gespaltene Stichproben (SSIV) zu arbeiten. Die erste und zweite Stufe werden mit unterschiedlichen Beobachtungen geschätzt. Die Verzerrungen tendieren dann nicht gegen den Verzerrungsgrad des OLS-Schätzers, sondern gegen Null. Vorliegende SSIV-Schätzungen zeigen jedoch zum Teil ausgesprochen unpräzise Ergebnisse, die je nach Datensatz und der Verwendung von speziellen Variablen je nach inhaltlicher Fragestellung schwanken.

Die Hahn-Hausman-Testprozedur ist, wie das Ausgangssystem (9) und (10) zeigt, nur für einen sehr einfachen Fall ausgelegt. Sind in (9) vorherbestimmte Variablen $x_1, ..., x_K$ enthalten, dann kann der Einfluss dieser Größen auf y_1 und y_2 vorher auspartialisiert werden. Das Vorgehen lässt sich verallgemeinern, wenn weitere gemeinsam abhängige Variablen existieren. Der Vorteil des Hahn-Hausman-Vorschlages zum Aufdecken schwacher Instrumente liegt gegenüber anderen Vorgehensweisen darin, dass das Verfahren abgestuft ist. Die Autoren zeigen anhand des Beispiels einer Nachfragefunktion, dass übliche R^2- und F-Werte bei einer großen Anzahl an Instrumenten nicht auf schwache Instrumente hindeuten, wohl aber der Hahn-Hausman-Test T_1. Ausgangs- und Umkehrregressionsschätzer nach 2SLS, aber auch nach LIML unterschieden sich in dem Beispiel deutlich, so dass weder der 2SLS- noch der LIML-Schätzer zu empfehlen ist. Durch Simulationsstudien verdeutlichen die Autoren, dass der 2SLS-Ausgangsschätzer in vielen Situationen, insbesondere wenn $l > 30$, ungeeignet ist, dass bei dem B2SLS-Schätzer keine Tendenz zu großen Teststatistiken besteht, dass aber auch der LIML-Schätzer gelegentlich im Mittel einen großen Bias aufweist.

3.4 Empirische Studien

Hauptanwendungsgebiet der IV-Schätzer bei mikroökonometrischen Untersuchungen ist bisher die Bestimmung der Ertragsraten der Schulbildung. Einen aktuellen Überblick über methodische Probleme und empirische Ergebnisse zu IV-Schätzungen liefert Card (2001). Ausgangspunkt zu einer ganze Reihe von Nachfolgeuntersuchungen ist die Studie von Angrist und Krueger (1991), in der erstmals mit dem Quartal des Geburtsjahres als Instrument für die Schuljahre gearbeitet wird. Weiterführende Analysen mit alternativen Instrumenten und für anderen Perioden sowie für andere Länder zeigen ein breites Spektrum an Ertragsraten der Schulbildung. Die durch IV-Schätzer ermittelten Ertragsraten liegen systematisch über denen der OLS-Schätzer, oft mehr als 20 Prozent. Dies steht im Einklang mit nach unten verzerrten OLS-Schätzern bei Fehlern in den exogenen Variablen. Andere Erklärungen können auch die folgenden sein: unbeobachtete Differenzen in den Untersuchungs- und Kontrollgruppen, eine systematisch verzerrende Auswahl von Instrumenten, die zu zu hohen Ertragsraten führen, die ausgewählten Instrumente beeinflussen in ihrem Verhalten vor allem oder ausschließlich diejenigen, die hohe Ertragsraten aufweisen.

4 Parametrische Paneldatenmodelle

Paneldaten gehören heute zum Alltag der empirischen Wirtschaftsforschung. Sie ermöglichen die Berücksichtigung individueller unbeobachteter Heterogenität, können zwischen Zustandsabhängigkeit und Heterogenität unterscheiden, können Perioden-, Alters- und Kohorteneffekte bestimmen. Paneldaten sind informativer als aggregierte Zeitreihendaten auf der einen Seite und Querschnittsdaten auf der anderen Seite. Sie weisen meist einen geringeren Grad an Multikollinearität und einen höheren Grad an Variabilität auf. Die Zahl der Freiheitsgrade liegt höher. Dies führt zu effizienteren Schätzungen. Die Dynamik lässt sich besser oder überhaupt erst durch Paneldaten erfassen. Die Effekte von Maßnahmen sind genauer zu identifizieren. Eine umfassendere Modellierung und Überprüfung individueller Verhaltsweisen ist möglich. Verzerrungen, die durch Aggregation hervorgerufen werden, unterbleiben.

Die traditionellen Paneldatenmodelle sind linear und unterscheiden zwischen fixen und zufälligen Individualeffekten. Im ersten Fall sind Korrelationen mit den Regressoren zugelassen. Außer der Zeitinvarianz werden keine weiteren Annahmen über den Individualeffekt gemacht. Im zweiten Fall ist der Individualeffekt als Teil der Störgröße modelliert, wobei die Verteilung des gesamten Störterms als bekannt und vollständig parametrisiert vorausgesetzt wird. Lineare Paneldatenmodelle sind gut erforscht, auch wenn ständig einzelne Aspekte vertieft untersucht und neue hinzugefügt werden. Zu nennen sind hier insbesondere dynamische Modellierungen, die Zeitreiheneigenschaften berücksichtigen, Fehler-in-den Variablen, simultane und multiplikative Effekte. Daneben gewinnen Linked-Employer-Employee-Modelle, lineare Mixed-Effect-Modelle und Mehrebenenanalysen an Bedeutung. Nichtlineare Paneldatenmodelle schließen insbesondere die Discrete-Choice-Modelle, zensierte, Selektions- und Zähldaten-Modelle ein. Neuere Überblicksbeiträge zur Paneldatenökonometrie liefern Arellano/Honore (2001), Arellano (2002), Baltagi (2001) und Lechner (2002).

4.1 Lineare Ansätze

Ausgangspunkt bei Paneldaten bildet das Modell

$$y_{it} = x'_{it}\beta + \alpha_i + \epsilon_{it} \qquad i = 1, \cdots, N; t = 1, \cdots, T \tag{14}$$

mit dem zeitinvarianten Individualeffekt α_i und einem klassischen Störterm ϵ_{it}. Die Individuen werden durch den Index i und die Perioden durch t erfasst. Einfache Erweiterungen folgen durch Berücksichtigung von Zeiteffekten, durch Einbeziehung von Autokorrelation und Heteroskedastie, verzögerte endogene Variable und Fehler in den Variablen (Hübler 1990). Hier kommen insbesondere ein-und zweistufige GMM-Schätzer zur Anwendung, die durch robuste Schätzer (Janz 1997) verbessert worden sind. Auch im Rahmen der Teststatistiken bei GMM-Ansätzen wurden Fortschritte erzielt vor dem Hintergrund, dass in kleinen Stichproben die asymptotischen Standardfehler auf Basis des effizienten zweistufigen GMM-Schätzers unterschätzt werden. Bond/Bowsher/Windmeijer (2001) z.B. schlagen kriteriumsbasierte Tests vor, die sich aufgrund von Simulationen gegenüber konventionellen Wald-Tests als überlegen zeigen. Oder Stock und Wright (2000) entwickeln eine asymptotische Schätz- und Testtheorie für den Fall schwach identifizierbarer Parameter , bei dem die üblichen GMM-Verfahren versagen.

Bei entsprechender Modifikation lassen sich die aus der Zeitreihenanalyse bekannten Methoden zum Aufdecken von Unit-Roots und Kointegration übertragen (Breitung 1992, 2000). Paneleffekte sind auch durch variable Koeffizienten modellierbar und im Rahmen interdependenter Modelle zu untersuchen (Hsiao 1986, Baltagi 2001). Im, Ahn, Schmidt und Wooldridge (1999) gehen der Frage nach, welche Schätzer bei strikt exogenen Variablen im Rahmen von Paneldatenmodellen effizient sind und zeigen, dass bei unbeschränkter Kovarianzmatrix GLS-Schätzer effizient sind, während bei Korrelation zwischen Regressoren und Störgrößen der 3SLS-Schätzer vorzuziehen ist. Für den Fall, dass beide Annahmen vorliegen, entwickeln die Autoren einen verallgemeinerten Instrumentalvariablen-Schätzer, der vom 3SLS-Schätzer abweicht und in modifizierter Form die Abweichungen vom Mittelwert berücksichtigt. Einen transformierten Likelihoodansatz schlagen Hsiao/Pesaran/Tahmiscioglu (2002) vor zur Schätzung fixer Effekte im dynamischen Paneldatenmodell bei einer geringen Zahl an Perioden.

4.1.1 Linked-Employer-Employee-Modelle

Die Verknüpfung von Individualdaten und Betriebs- bzw. Unternehmensdaten stellt eine neue Informationsquelle dar, aus der gegenüber bisherigen Datenquellen eine Reihe zusätzlicher Erkenntnisse gezogen werden können. Diese heute allgemein unter dem Begriff "linked employer-employee (LEE) data " zusammengefasste Form der Daten umfasst repräsentative betriebliche Querschnittsdaten mit repräsentativen oder nichtrepräsentativen Informationen der Arbeitskräfte, repräsentative Querschnittsinformation der Beschäftigten mit zugespielten Längsschnittdaten der Betriebe als auch gematchte Betriebs-und Individualpaneldaten. Insbesondere LEE-Paneldaten (LEEP) bieten neue Möglichkeiten der Auswertung. Zu nennen sind hier vor allem das Aufdecken von Erhebungsfehlern bei einer der beiden Datenquellen (Betriebe, Individuen). Es lassen sich inner- und zwischenbetriebliche Effekte unterscheiden. Die relative Bedeutung von Belegschaftsmerkmalen auf der einen Seite und Unternehmenscharakteristika auf der anderen Seite kann ermittelt werden und Interaktionseffekte müssen nicht mehr unerkannt bleiben. Es ist möglich zu ermitteln, ob unbeobachtete Heterogenität eher personen-oder eher betriebsbedingt ist. Entscheidungen der Unternehmer und Manager bewirken Veränderungen im Verhalten der Beschäftigten und umgekehrt. Diese Effekte können in einem dynamischen LEEP-Modell bestimmt werden. Eine ausführliche Diskussion zu den Möglichkeiten, LEE-Daten auszuwerten, führt Hamermesh (1999).

Auch aus methodischer Sicht sind verschiedene Aspekte des LEEP-Modells von Interesse. Einen Überblick hierzu liefern Abowd/Kramarz (1999, 1999a), Abowd, Kramarz und Margolis (1999) und Goux/Maurin (1999). Das LEEP-Grundmodell ist eine Verallgemeinerung des traditionellen Paneldatenmodells (14) mit fixen Effekten. Neben dem Individualeffekt α_i wird ein Firmeneffekt (Betriebseffekt)ψ_J berücksichtigt

$$y_{it} = x'_{it}\beta + \alpha_i + \psi_{J(it)} + \epsilon_{it}, \tag{15}$$

wobei $i = 1, \cdots, N$ die Personen, $j = 1, \cdots, J$ die Betriebe und $t = 1, \cdots, T$ die Perioden beschreiben. Für den Störterm ϵ_{it} gilt

$$E[\epsilon_{it}|i, J(it), t, x_{it}] = 0$$

und

$$Cov[\epsilon_{it}\epsilon_{i'\tau}|i, i', J(it), J(i'\tau), t, \tau, x_{it}, x_{i'\tau}] = \begin{cases} \sigma_\epsilon^2 & wenn \quad i = i' \quad und \quad t = \tau \\ 0 & sonst. \end{cases}$$

In Matrixschreibweise entspricht (15)

$$y = X\beta + D\alpha + F\psi + \epsilon, \tag{16}$$

wobei y ein $N \cdot T \times 1$-Vektor, X eine $N \cdot T \times K$-Matrix, D eine $N \cdot T \times N$-Matrix von (0;1)-Indikatoren für N Beschäftigte, F eine $N \cdot T \times J$-Matrix von (0;1)-Indikatoren für J Betriebe, in denen die N Personen in T Perioden arbeiten. Als Erweiterung wird vorgeschlagen (Abowd/Kramarz/Margolis 1999, 265), die Individual- und Firmeneffekte nicht nur als unbeobachtete Komponente zu modellieren, sondern auch systematisch variierende Veränderungen zuzulassen. Gedacht ist im ersten Fall an zeitinvariante Personencharakteristika und im zweiten Fall vor allem an die individuelle Betriebszugehörigkeitsdauer.

Das Anliegen von Modell (16) ist, die Koeffizienten β, α und ψ zu bestimmen, insbesondere die unbeobachteten Individual- und Betriebseffekte. Die Matrix der Firmeneffekte F ist orthogonal zum Störgrößenvektor ϵ. Es wird damit "exogenous mobility" unterstellt. Existieren Betriebseffekte und bleiben diese wie im traditionellen Paneldatenmodell für Indidividuen unberücksichtigt, so sind die OLS-Schätzungen für α und β verzerrt($\hat{\alpha}^*, \hat{\beta}^*$). Dies ergibt sich z.B. sofort aus der Erwartungswertbildung der Schätzfunktion des Parametervektors aus dem partitionierten Modell (14), wenn aber (15) das korrekte Modell ist, d.h.

$$E[\hat{\beta}^*] = E[(X'P_D X)^{-1}X'P_D(X\beta + D\alpha + F\psi + \epsilon)] = E[\beta + (X'P_D X)^{-1}XP_D F\psi] \neq \beta$$

mit $P_D = I - D(D'D)^{-1}D'$ (vgl. z.B. Hübler 1989, 46). Anstelle von β in (14) wird hier β^* als Symbol verwendet, um den Unterschied zu β in (15) deutlich zu machen. Ähnliches ergibt sich für den Individualeffekt

$$E(\hat{\alpha}^*) = \alpha + (D'P_X D)^{-1}D'P_X F\Psi$$

Der Verzerrungsgrad des Individualeffektes lässt sich als personendurchschnittlicher Firmeneffekt interpretieren

$$\alpha_i^* - \alpha_i = \sum_{t=1}^{T} \frac{\psi_{J(i,t)}}{T}.$$

Wenn nur das übliche Paneldatenmodell für Betriebe herangezogen, dann ist der Betriebseffekt genauso verzerrt geschätzt wie der Vektor β, wenn die Individualeeffekte außer Acht bleiben. Die Verzerrung ist dann der Durchschnitt der Personeneffekte.

Soll das "wahre" Modell (16) geschätzt werden, so ist dies praktisch wegen der Größenordnung nur schwer handhabbar. Einen direkten KQ-Algorithmus präsentieren Abowd/Creecy/Kramarz (2002). Ansonsten werden verschiedene approximative Lösungen vorgeschlagen. Hierzu wird das Gleichungssystem erweitert. Die Variablen Z, die aus Informationen von X, D und F zustande kommen, werden eingefügt, deren Effekte auf y sind durch λ bestimmt

$$(17)\qquad \begin{bmatrix} X'X & X'D & X'F \\ D'X & D'D & D'F \\ F'X & F'D & F'F \end{bmatrix} \begin{bmatrix} \beta \\ \alpha \\ \psi \end{bmatrix} = \begin{bmatrix} X'y \\ D'y \\ F'y \end{bmatrix}.$$

Vorgeschlagen werden verschiedene approximative Lösungen. Hierzu wird das Gleichungssystem erweitert. Die Variablen Z, die aus Informationen von X, D und F zustande kommen, werden eingefügt, deren Effekte auf y sind durch λ bestimmt

$$(18)\qquad y = X\beta + D\alpha + Z\lambda + P_Z F\psi + \epsilon =: \tilde{X}\tilde{\beta} + \epsilon.$$

Zur Schätzung der Kovarianzmatrix $\hat{V}(\hat{\tilde{\beta}}) = \hat{\sigma}_\epsilon^2(\tilde{X}'\tilde{X})^{-1}$ sind die Residuen

$$\hat{\epsilon} = y - X\hat{\beta} - D\hat{\alpha} - Z\hat{\lambda} - P_Z F\hat{\psi}$$

zu bestimmen. Dies erfolgt in drei Schritten

(i) $\hat{\epsilon}^{(1)} = (y - X\hat{\beta}) - D\hat{\alpha} - Z\hat{\lambda}$

(ii) $\hat{\epsilon}^{(2)} = P_Z F\hat{\psi} = F\hat{\psi} - Z\hat{\lambda}$

(iii) $\hat{\epsilon} = \hat{\epsilon}^{(1)} - \hat{\epsilon}^{(2)}$.

Daraus kann

$$\hat{\sigma}_\epsilon^2 = \frac{1}{NT - K - N - J + 1}\hat{\epsilon}'\hat{\epsilon}$$

errechnet werden. Spezielle Modellierungen der Individual- und Firmeneffekte nehmen Goux und Maurin (1999, 524ff) vor. Sie lassen unter anderem Interaktionen zwischen Betriebs- und Individualeffekten sowie zwischen Personen- und Firmencharakteristika zu und führen zufällige Sektoreffekte ein.

Die neu strukturierte Gleichung (18) ist identisch mit (16), denn der Hilfsparametervektor λ ist definiert als

$$\lambda = (Z'Z)^{-1}Z'F\psi.$$

Somit gilt

$$Z\lambda + P_Z F\psi = Z(Z'Z)^{-1}Z'F\psi + (I - Z(Z'Z)^{-1}Z')F\psi = F\psi.$$

Wenn zwei bedingte Orthogonalitätsnahmen gemacht werden, nämlich

- X und F sind orthogonal bei gegebenem Z
- D und F sind orthogonal bei gegebenem Z

dann lässt sich das erweiterte System (18) einfacher als das Ausgangssystem (16) schätzen, da ein blockdiagonales System vorliegt

$$\begin{bmatrix} \hat{\beta} \\ \hat{\alpha} \\ \hat{\lambda} \\ \hat{\psi} \end{bmatrix} = \begin{bmatrix} X'X & X'D & X'Z & 0 \\ D'X & D'D & D'Z & 0 \\ Z'X & Z'D & Z'Z & 0 \\ 0 & 0 & 0 & F'P_Z F \end{bmatrix}^{-} \begin{bmatrix} X'y \\ D'y \\ Z'y \\ F'P_Z y \end{bmatrix}.$$

denn es gilt außerdem $Z'P_Z F = 0$. Statt den ersten Block $\hat{\beta}, \hat{\alpha}, \hat{\lambda}$ getrennt von $\hat{\psi}$ zu schätzen, kann äquivalent auch vierstufig verfahren werden.

(i) Schätze

$$\begin{bmatrix} \hat{\beta} \\ \hat{\lambda} \end{bmatrix} \begin{bmatrix} X'P_D X & X'P_D Z \\ Z'P_D X & Z'P_D Z \end{bmatrix}^{-1} \begin{bmatrix} X'P_D y \\ Z'P_D y \end{bmatrix}.$$

(ii) Bestimme unter Verwendung von $\hat{\beta}$ und $\hat{\lambda}$

$$\hat{\alpha} = (D'D)^{-1} D'(y - X\hat{\beta} - Z\hat{\lambda}).$$

(iii) Bilde die partitionierte Hilfsregression $y = F\psi + Z\pi + v$ und ermittle

$$\hat{\pi} = (Z'P_F Z)^{-1} Z'P_F y.$$

(iv) Berechne unter Zuhilfenahme von $\hat{\pi}$

$$\hat{\psi} = (F'F)^{-1} F'(y - Z\hat{\pi}).$$

Wird Modell (15) um (Wirtschafts-)Sektoreffekte (Industrieeffekte) erweitert, dann ist zu beachten, dass sich die Sektoreffekte κ_l mit $l = 1, \cdots, L$ als beschäftigungsgewichteter Durchschnitt der Firmeneffekte darstellen lassen. Jeder Betrieb ist einem Wirtschaftssektor zuzuordnen, so dass (15) übergeht in

$$y_{it} = x'_{it}\beta + \alpha_i + \kappa_l(J(it)) + (\psi_{J(it)} - \kappa_l(J(it)) + \epsilon_{it} \tag{19}$$

bzw. bei Matrixschreibweise in

$$y = X\beta + D\alpha + FA\kappa + (F\psi - FA\kappa) + \epsilon. \tag{20}$$

Die Matrix A nimmt die Zuordnung eines Betriebes zu einem Sektor vor ($a_{jl} = 1$, wenn Betrieb j zu Sektor l gehört; $a_{jl} = 0$ sonst). Der Betriebseffekt in Abhängigkeit des Sektoreffektes geschätzt, d.h. aus

$$F\psi = FA\kappa + \omega,$$

ergibt

$$\hat{\kappa} = (A'F'FA)^{-1}A'F'F\psi.$$

Damit entspricht

$$F\psi - FA\hat{\kappa} = F\psi - FA(A'F'FA)^{-1}A'F'F\psi = (I - FA(A'F'FA)^{-1}A'F')F\psi =: P_{FA}F\psi.$$

Wenn die Sektoreffekte nicht über (20), sondern unter Ausschluss der Betriebseffekte geschätzt werden, also $P_{FA}F\psi$ in den Störterm geht, dann folgt der reine Sektoreffekt einschließlich einem Bias, der sich als gewichteter Firmeneffekt beschreiben lässt. Dieser Bias entfällt nur, wenn FA und $P_{FA}F$ orthogonal sind. Wenn außerdem der Individualeffekt unberücksichtigt bleibt, dann entspricht der geschätzte (rohe) Sektoreffekt aus

$$y = X\beta + FA\kappa + \epsilon$$

einem gewichteten Durchschnitt aus Individual- und Firmeneffekt

(21) $$\kappa^* = (A'F'P_XFA)^{-1}A'F'P_XD\alpha + (A'F'P_XFA)^{-1}A'F'P_XF\psi.$$

Analog zu den Sektoreffekten ist es möglich, Berufseffekte in einem erweiterten LEEP-Modellrahmen zu untersuchen, die häufig als Interaktion aus Individual- und Betriebseffekt interpretiert werden (Groshen 1996). Weitere Analysemöglichkeiten der LEEP-Modelle ergeben sich gegenüber reinen Panelmodellen mit Betriebsdaten, da Betriebseffekte unter Berücksichtigung unbeobachteter Individualeffekte extrahiert werden können. Hildreth und Pudney (1999) gehen der Frage nach, welche Probleme sich bei nichtzufallsbedingten fehlenden Werten in LEEP-Modellen ergeben. Endogenität und Messfehler bei den Regressoren führen auch bei LEEP-Modellen zur Anwendung von IV-Schätzern, worauf Goux und Maurin (1999) aufmerksam machen.

4.1.2 Mixed-Effects-Modelle

Mixed-Effects-Modelle oder kurz nur Mixed-Modelle (Searle 1971, 380ff., Verbeke und Molenberghs 2000) enthalten fixe und zufällige Effekte, sind im Sinne von Paneldatenmodellen jedoch reine Random-Effects-Modelle. Ausgegangen wird von

(22) $$y_i = X_i\beta + Z_i\beta_i + u_i,$$

wobei y_i ein $N_i \times 1$-Vektor ist. Falls N_i für alle i gleich der Zahl der Perioden T entspricht, liegt der Paneldatendatenfall vor. Hiervon wird im Folgenden ausgegangen, so dass im Balanced-Panel gilt: $X_i \sim T \times K$ und $Z_i \sim T \times M$. Das einfache RE-Modell ergibt sich, wenn $Z_i = \iota_i \sim T \times 1$ und $\beta_i \sim 1 \times 1$. Eine Zusammenfassung von Individual- und Betriebseffekten ergibt sich, wenn

$$\begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & & & & \\ 1 & 0 & 0 & \cdots & 1 \end{pmatrix}_{T\times(J+1)}.$$

Der Vektor β ist ein unbekannter fixer $K \times 1$-Vektor und β_i ein $(M \times 1)$-Vektor von unbeobachteten Zufallseffekten. Für die Zufallsvariablen wird unterstellt, dass sie untereinander und von X unabhängig sind ($E(\beta_i|X) = 0; E(u_i|X) = 0$). Außerdem gilt

$$\beta_i \sim N(0;\Omega) \quad und \quad u_i \sim N(0;\Sigma_i), \tag{23}$$

so dass

$$y_i \sim N(X_i\beta; Z_i\Omega Z_i' + \Sigma_i). \tag{24}$$

Falls $X_i = Z_i$, dann ist der Mixed-Effects-Ansatz ein Modell mit zufällig variierenden Koeffizienten, d.h. der Koeffizientenvektor $\ddot{\beta} = \beta + \beta_i$ besteht aus einem fixen und einem zufällig variierenden Teil (Longford 1993, 95). Als übliches Schätzverfahren für β wird der ML-Ansatz gewählt, der dem iterativen GLS-Schätzer entspricht, wobei zunächst von der Kenntnis des Parametervektors α aus der Kovarianzmatrix $V_i = Z_i\Omega Z_i' + \Sigma_i$ ausgegangen wird, in dem alle Varianz- und Kovarianzparameter zusammengefasst sind.

$$\hat{\beta}(\gamma) = (\sum_{i=1}^{N} X_i'V_i(\alpha)^{-1}X_i)^{-1}\sum_{i=1}^{N} X_i'V_i(\gamma)^{-1}y_i. \tag{25}$$

Zur Schätzung von α kann die restringierte ML-Schätzung (REML) herangezogen werden, d.h. maximiert wird

$$L_{REML}(\beta;\alpha) = |(\sum_{i=1}^{N} X_i'V_i^{-1}(\alpha)X_i)|^{-\frac{1}{2}}L_{ML}(\beta;\alpha), \tag{26}$$

wobei nach α und β abzuleiten ist. Zusammengefasst über alle Beobachtungen ($i = 1,\cdots,N$) bei Unabhängigkeit zwischen den Perioden, so dass

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 & \cdots & 0 \\ 0 & \Sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_N \end{bmatrix} \tag{27}$$

und

$$\Omega^g = \begin{bmatrix} \Omega & 0 & \cdots & 0 \\ 0 & \Omega & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Omega \end{bmatrix}, \tag{28}$$

ergeben sich die Henderson-Mixed-Modellgleichungen (Verbeke/Molenberghs 2000, 79f)

$$\begin{pmatrix} X'\Sigma^{-1}X & X'\Sigma^{-1}Z \\ Z'\Sigma^{-1}X & Z'\Sigma^{-1}Z + \Omega^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ \beta^g \end{pmatrix} = \begin{pmatrix} X'\Sigma^{-1}y \\ Z'\Sigma^{-1}y \end{pmatrix}, \tag{29}$$

wobei $\beta^g = (\beta_1,\beta_2,\cdots,\beta_M)'$. Die Lösung dieser Gleichungen ist bei großen Datensätzen rechentechnisch nur schwer zu bewältigen. Praktisch ist (25) zu präferieren. Einen zweistufigen Schätzer schlagen Yang, Yu und Al-Zaid (2001) vor. Dies hat den Vorteil, dass auf die Normalverteilungsannahme verzichtet werden kann. Zudem steht man nicht vor den praktischen

Rechenproblemen wie beim REML-Ansatz. Zunächst werden β, β_i und σ^2 für jeden Beobachtungsträger getrennt geschätzt ($\hat{\beta}_{(i)}, \hat{\beta}_i, \hat{\sigma}_i^2$). Anschließend sind die arithmetischen Mittel zu bilden ($\bar{\beta}, \bar{\beta}^g, \bar{\sigma}^2$), die unverzerrte Schätzer für β, β^g und σ^2 darstellen. Bei $\bar{\sigma}^2$ ist auf die Freiheitsgrade zu achten

$$(30) \qquad \bar{\sigma}^2 = \frac{T-K-M}{NT-N(K+M)} \sum_{i=1}^{N} \hat{\sigma}_i^2.$$

Die Kovarianzmatrix Ω folgt dann mit

$$(31) \qquad \hat{\Omega} = \frac{1}{N-1} \sum_{i=1}^{N} (\hat{\beta}_i - \bar{\beta}^g)(\hat{\beta}_i - \bar{\beta}^g)' - \frac{\bar{\sigma}^2}{N} \sum_{i=1}^{N} (Z_i' P_{X_i} Z_i)^{-1},$$

wobei $P_{X_i} = I - X_i(X_i'X_i)^{-1}X_i'$. Daraus kann

$$(32) \qquad \hat{V}(\bar{\beta}) = \frac{1}{N^2} \sum_{i=1}^{N} \hat{V}(\hat{\beta}_(i)) = \frac{1}{N^2} \sum_{i=1}^{N} \bar{\sigma}^2 (X_i' P_{Z_i} X_i)^{-1}$$

und

$$(33) \qquad \hat{V}(\bar{\beta}^g) = \frac{1}{N^2} \sum_{i=1}^{N} \hat{V}(\hat{\beta}_i) = \frac{1}{N^2} \sum_{i=1}^{N} \hat{\Omega} + (Z_i' P_{X_i} Z_i)^{-1}$$

mit $P_{Z_i} = I - Z_i(Z_i'Z_i)^{-1}Z_i$ ermittelt werden. Anstelle der einfachen KQ-Schätzung ($\bar{\beta}, \bar{\beta}^g$) lässt sich eine gewichtete KQ-Schätzung durchführen, bei der

$$(34) \qquad \bar{\beta}_W = [\sum_{i=1}^{N} \hat{V}(\hat{\beta}_(i))^{-1}]^{-1} [\sum_{i=1}^{N} \hat{V}(\hat{\beta}_(i))^{-1} \hat{\beta}_(i)]$$

gebildet wird. Analog ist $\bar{\beta}_W^g$ zu bestimmen. Simulationsstudien zeigen, dass die gewichteten KQ-Schätzungen besser abschneiden als die einfachen KQ-Schätzungen, dass bei der Varianzschätzung das zweistufige Verfahren nur geringfügig dem REML-Ansatz unterlegen ist. Insgesamt kann das Mixed-Effects-Modell als Alternative zum Fixed-Effects-Ansatz der Linked-Employer-Employee-Modelle des vorangegangen Abschnitts betrachtet werden. Individual- und Firmeneffekte - vgl. (15) - sind dann als Zufallseffekte aufzufassen, d.h.

$$(35) \qquad \begin{pmatrix} \alpha \\ \psi \end{pmatrix} \sim N(0; \Omega).$$

Für die Ω-Matrix können Korrelationen zwischen Individual- und Betriebseffekten zugelassen werden. Es liegt dann keine blockdiagonale Matrix vom Typ (28) mehr vor. Wenn neben zufälligen Individual- und Betriebseffekten noch zufällige Sektoreffekte modelliert werden, dann entsteht ein Vierebenenmodell. In der einfachsten Form entspricht dies folgendem Random-Effects-Modell (i=1,...,N - Individuen; j=1,...,J - Betriebe; l=1,...,L - Sektoren; t=1,...,T - Perioden)

$$(36) \qquad y_{ijlt} = x_{ijlt}'\beta + \alpha_i + \alpha_{ij} + \alpha_{ijl} + \epsilon_{ijlt}$$

mit den vier unabhängigen Zufallsvariablen $\alpha_i \sim N(0; \sigma_4^2)$, $\alpha_{ij} \sim N(0; \sigma_3^2)$, $\alpha_{ijl} \sim N(0; \sigma_2^2)$, $\epsilon_{ijlt} \sim N(0; \sigma_1^2)$, so dass

$$V(y_{ijlt}) = \sigma_4^2 + \sigma_3^2 + \sigma_2^2 + \sigma_1^2.$$

Je niedriger die Ebene ist (Ebene 1: Personen, Ebene 2: Betriebe, Ebene 3: Sektoren, Ebene 4: Perioden), auf der Variationen zugelassen werden, aus um so mehr Elementen bestehen die Kovarianzen

$$Cov(y_{ijlt}, y_{i'jlt}) = \sigma_4^2 + \sigma_3^2 + \sigma_2^2$$

$$Cov(y_{ijlt}, y_{i'j'lt}) = \sigma_4^2 + \sigma_3^2$$

$$Cov(y_{ijlt}, y_{i'j'l't}) = \sigma_4^2.$$

Goldstein (1995) und Longford (1993, S.157ff) diskutieren ausführlich die Spezifikation und Schätzung von Mehrebenenmodellen.

Das Problem der Random-Effects-Modelle (REM) ist, dass sie bei Hausman-Tests gegenüber Fixed-Effects-Modellen (FEM) üblicherweise abgelehnt werden, da Abhängigkeiten zwischen Regressoren und Störterm bestehen. Fixed-Effects-Schätzer, wie sie bei Linked-Employer-Employee-Ansätzen bevorzugt werden, lassen keine Berücksichtigung von Dummy-Variablen zu. Sie entfallen bei diesem Vorgehen aufgrund der Differenzenbildung als Regressoren. Naheliegend ist daher eine Kombination aus beiden Ansätzen zu entwickeln. Ausgangspunkt bildet die vereinfachte Version eines Zweiebenenmodells

(37) $$y_{ij} = x'_{ij}\beta + \alpha_j + \epsilon_{ij} = x'_{ij}\beta + u_{ij}$$

mit $i = 1, \cdots, N$ als Zahl der Beobachtungsträger auf der ersten Ebene (Personen) und $j = 1, \cdots, J$ als Zahl der Beobachtungsträger auf der zweiten Ebene (Betriebe). Falls die Abhängigkeit zwischen Regressoren und Störgröße allein auf x_{ij} und α_j zurückzuführen ist, so kann der Erwartungswert des Teilstörterms α_j explizit als deterministische Größe modelliert werden. Der Gedanke hierzu findet sich beim Sample-Selection-Ansatz. Damit entfällt das Problem der Abhängigkeit zwischen Störterm und Regressoren. Es muss aber auch nicht wie beim FEM über Differenzenbildung oder Within-Modellierung zur Schätzung von β der Betriebseffekt entfernt werden, wodurch auch [0;1]-Regressoren als Determinanten verschwinden. Bei dieser Art des Vorgehens ist iterativ in folgenden Schritten zu verfahren:

- Ziel ist das Modell in (22) so zu transformieren, dass asymptotisch ein klassisches Regressionsmodell entsteht.
- Ermittle eine konsistente Schätzung für den Erwartungswert von α_j und erweitere die Ausgangsgleichung (22) um einen Term, der ein unbekanntes Vielfaches (β_j) der Schätzung von $E(\alpha_j)$ beschreibt:

  $$\hat{\alpha}_j \cdot \delta_j = \hat{\bar{u}}_j^{(1)} \cdot \delta_j.$$

  Eine konsistente Schätzung lässt sich wegen der Abhängigkeit von x_{ij} und α_j nicht über eine RE-, aber über eine FE-Schätzung erzielen.
- Schätze β im erweiterten Modell

  $$y_{ij} = x'_{ij}\beta + \hat{\bar{u}}_j^{(1)}\delta_j + \tilde{u}_{ij}$$

  und bilde daraus erneut das durchschnittliche Residuum $\hat{\bar{u}}_j^{(2)} = (1/N_j)\sum_{i=1}^{N_j} \hat{\bar{u}}_{ij}$ für j.

- Revidiere das erweiterte Modell, indem $\hat{u}_j^{(1)}$ durch $\hat{u}_j^{(2)}$ ersetzt wird.

Auf diese Weise kann weiter verfahren werden, bis $\hat{u}_j^{(s-1)} \approx \hat{u}_j^{(s)}$, bis der geschätzte Koeffizient $\hat{\delta}^{(s)}$ gegen Eins tendiert. Alternativ kann Modell (37) als partiell linearer Ansatz aufgefasst werden und der Individualterm nichtparametrisch geschätzt werden (vgl. 5.1.2)

Im Einzelnen ist bei diesem iterativen Verfahren wie folgt zu verfahren. In Matrixschreibweise ist das FEM über alle Betrieb j=1,...,J

$$y = X\beta + D\alpha + \epsilon \tag{38}$$

wobei $\alpha = (\alpha_1, ..., \alpha_J)'$, $\iota' = (1, ..., 1)$ und

$$D = \begin{pmatrix} \iota_1 & & \\ & \ddots & \\ & & \iota_J \end{pmatrix}.$$

Für den Beschäftigten i im Betrieb j reduziert sich dieser Zusammenhang auf

$$y_{ij} = x'_{ij}\beta + d'_j\alpha + \epsilon_{ij} \tag{39}$$

mit $d'_j = (0, \cdots, 0, 1, 0, \cdots, 0)_{1\times J}$. Wird

$$Q_j = I_{N_j} - \frac{1}{N_j}\iota_{N_j}\iota'_{N_j} = I_{N_j} - \bar{E}_j,$$

definiert, wobei $\iota'_{N_l} = (1, ..., 1) \sim N_l \times 1$, so kann durch Prämultiplikation von (38) mit Q_J die Gleichung für den Within-Schätzer gebildet werden

$$Q_l y_j = Q_l X_j\beta + Q_j D_j\alpha + Q_j\epsilon_j = Q_j X_j\beta + Q_j\epsilon_j. \tag{40}$$

Aggregiert über die Beobachtungen aller Betriebe mit

$$\begin{pmatrix} Q_1 & & \\ & \ddots & \\ & & Q_J \end{pmatrix} \tag{41}$$

und

$$\begin{pmatrix} X_1 \\ \vdots \\ X_L \end{pmatrix} \tag{42}$$

folgt als Schätzer

$$\hat{\beta}_W = (X'Q'QX)^{-1}X'Q'Qy. \tag{43}$$

Hieraus wird der Durchschnittsresiduenvektor über alle Betriebe

$$\hat{u}' = (\hat{u}_1, ..., \hat{u}_J) = (\bar{y} - \bar{X}\hat{\beta}_W)' \tag{44}$$

gebildet. Als erweitertes Modell folgt

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_J \end{pmatrix} = X\beta + \begin{pmatrix} \hat{\bar{u}}_1 \iota_{N_1} & & \\ & \ddots & \\ & & \hat{\bar{u}}_J \iota_{N_J} \end{pmatrix} \begin{pmatrix} \delta_1 \\ \vdots \\ \delta_J \end{pmatrix} + \tilde{\epsilon} = \tag{45}$$

$$X\beta + diag(\hat{\bar{u}}_1 \iota_{N_1}, \cdots, \hat{\bar{u}}_J \iota_{N_J})\delta + \tilde{u} =: X\beta + \Delta\delta + \tilde{\epsilon} = X\beta + \tilde{u}.$$

Der OLS-Schätzer aus dem partitionierten Modell für β ist dann

$$\hat{\beta}^{(1)}_{OLS} = (X'P_\Delta X)^{-1} X' P_\Delta y. \tag{46}$$

Jetzt können erneut die durchschnittlichen Residuen je Betrieb ermittelt und danach in (45) die bisherigen Residuen ersetzt werden. Anschließend lässt sich ein weiterer Iterationsschritt durchführen. Das Modell (45) entspricht asymptotisch der transformierten Gleichung (40), die der Within-Schätzung des FEM dient. Da letztere konsistent ist, muss dies auch für die (46) der Fall sein.

4.1.3 Empirische Studien

Linked-Employer-Employee-Modelle werden bisher überwiegend zur Bestimmung von Lohneinkommenseffekten herangegezogen. Für Frankreich liegen die meisten LEEP-Studien vor, z.B. von Abowd, Kramarz, Margolis 1999, Goux, Maurin 1999. Mit niedersächsischen Gehalts- und Lohnstrukturerhebungsdaten hat Stephan (2001) gearbeitet. Einen Überblick über neuere Ergebnisse mit gematchten Employer-Employee-Daten liefern Abowd und Kramarz (1999). Eine Zusammenstellung von Arbeiten mit Daten der gematchten Datei aus dem IAB-Betriebspanel und der Beschäftigtenstatistik ist in Bellmann/Bender/Kölling (2000) zu finden. Mehrebenenmodelle sind vor allem bei Untersuchungen über die Effektivität von Schultypen eingesetzt worden (Goldstein 1995). Für regionale Lohnkurven unter Verwendung von Betriebspaneldaten schätzen Bellmann und Blien (2001) ein Dreiebenenmodell, das die Beschäftigten, Kreise und Bundesländer explizit erfasst.

4.2 Nichtlineare Ansätze

Das zentrale Problem nichtlinearer Paneldatenmodelle besteht gegenüber den linearen Ansätzen darin, dass durch einfache Differenzenbildung der Individualeffekt meist nicht zu entfernen ist. Für einzelne Modelltypen existieren spezielle Lösungen, die in breiten Bereichen Anwendung gefunden haben.

4.2.1 Fixed–Effects–Logit–ML–Schätzer

Bedingte Maximum-Likelihood-Schätzer lassen sich auf Logitmodelle mit fixen Individualeffekten anwenden (Hsiao 1986, 159ff). Das Grundmodell lautet

$$(47) \qquad y_{it}^* = x_{it}'\beta + \alpha_i + \epsilon_{it}$$

$$(48) \qquad y_{it} = \begin{cases} 1 & \text{, wenn} \quad y_{it}^* \geq 0 \\ 0 & \text{sonst} \end{cases}$$

$$(49) \qquad P(y_{it} = 1) = \frac{\exp\,(x_{it}'\beta + \alpha_i)}{1 + \exp\,(x_{it}'\beta + \alpha_i))}.$$

Eine einfache ML–Schätzung des Logit-Modells ist inkonsistent. Als Alternative bietet sich die CML–Schätzung (bedingte ML–Schätzung) an:

$$L^c = \prod_{i=1}^{N} P(Y_{i1} = y_{i1}, \ldots, Y_{iT} = y_{iT} | \sum_t y_{it})$$

$$P(y_i | \sum_{t=1}^{T} y_{it}) = \exp\,((\sum_{i=1}^{T} x_{it} y_{it})'\beta) / \sum_{d \epsilon \tilde{B}_i} \exp\,((\sum_{t=1}^{T} x_{it} d_{it})'\beta),$$

wobei $\tilde{B}_i = (d_{i1}, \ldots, d_{iT}) | d_{it}$ die Werte 0 oder 1 annimmt und $\sum_t d_{it} = \sum_t y_{it}$.

$\tilde{B}_i$ stellt alternative Datensätze bzgl. y_{it} dar, die die Bedingung $\sum_t y_{it}$ erfüllen. Bei T Beobachtungen kann die Zahl der Elemente mit dem Wert 1 folgenden Gesamtwert annehmen:

$$\sum_{t=1}^{T} y_{it} = 0, 1, \ldots, T.$$

$\sum y_{it} = 0$ und $\sum y_{it} = T$ tragen nichts zur Likelihood-Funktion bei, da es hierfür jeweils nur eine Kombination gibt und somit die Wahrscheinlichkeit 1 ist. Also ist nur die Summe von 1 bis T-1 relevant. Die bedingte Wahrscheinlichkeit für den Spezialfall T=2 ist dann

$$\begin{aligned} P(\omega_i = 1 | y_{i1} + y_{i2} = 1) &= \frac{P(\omega_i = 1)}{P(\omega_i = 1) + P(\omega_i = 0)} \\ &= \frac{[exp(x_{i2}'\beta)/(1 + exp(x_{i2}'\beta))] \cdot [exp(-x_{i1}'\beta)/(1 + exp(-x_{i1}'\beta))]}{\frac{exp(x_{i2}'\beta)}{1+exp(x_{i2}'\beta)} \cdot \frac{exp(-x_{i1}'\beta)}{1+exp(-x_{i1}'\beta)} + \frac{exp(x_{i1}'\beta)}{1+exp(x_{i1}'\beta)} \cdot \frac{exp(-x_{i2}'\beta)}{1+exp(-x_{i2}'\beta)}} \\ &= \frac{exp[(x_{i2} - x_{i1})'\beta]}{1 + exp[(x_{i2} - x_{i1})'\beta]} = F[(x_{i2} - x_{i1})'\beta] \end{aligned}$$

mit $\omega_i = 1$, wenn $(y_{i1}; y_{i2}) = (0; 1)$, und $\omega_i = 0$, wenn $(y_{i1}; y_{i2}) = (1; 0)$. Die bedingte Likelihood–Funktion

$$\ln L^c = \sum \{w_i \cdot \ln F[(x_{i2} - x_{i1})'\beta] + (1 - w_i) \cdot \ln[1 - F[(x_{i2} - x_{i1})'\beta]]\}$$

kann mit üblichen Mitteln wie bei einem einfachen Logit-Modell gelöst werden.

Statt einen Logit–Ansatz bei Vorliegen von fixen Effekten zu verwenden, kann auch ein einfaches Wahrscheinlichkeitsmodell benutzt werden, bei dem sich auch einfache Funktionen für die interessierenden Parameter, die unabhängig von den "nuisance"-Parametern sind, finden lassen. Für Probit–Modelle existieren keine Möglichkeiten dieser Art (Hsiao 1986, S. 163).

4.2.2 Random–Effects–Probit–ML–Schätzer

Ausgangspunkt bildet das Modell

$$(50) \qquad \begin{aligned} y_{it}^* &= x_{it}'\beta + \alpha_i + \epsilon_{it} \\ &= x_{it}'\beta + u_{it}; \qquad i = 1, \ldots, N \qquad t = 1, \ldots, T \end{aligned}$$

$$\begin{aligned} y_{it} &= \begin{cases} 1, & wenn \quad y_{it}^* \geq 0 \\ 0 & sonst \end{cases} \\ V(\alpha_i + \epsilon_{it}) &= V(u_{it}) = \sigma_\alpha^2 + \sigma_\epsilon^2 \\ \rho^2 &= Corr^2(u_{it}; u_{it'}) = \sigma_\alpha^2 / (\sigma_\alpha^2 + \sigma_\epsilon^2) \end{aligned}$$

$$P(Y_{i1} = y_{i1}, \cdots, Y_{iT} = y_{iT}) = \int_{-x_{i1}'\beta}^{\infty} \cdots \int_{-x_{iT}'\beta}^{\infty} f(u_{i1}, \cdots, u_{iT}) du_{iT} \cdots du_{i1}.$$

Hierauf lässt sich der CML-Schätzer nicht anwenden. Stattdessen kann der Butler-Moffitt-Schätzer unter Verwendung des Gaußchen Quadraturverfahrens herangezogen werden

$$(51) \qquad P(y_{i1}, \cdots, y_{iT}) = \int_{-x_{i1}'\beta}^{\infty} \cdots \int_{-x_{iT}'\beta}^{\infty} \int_{-\infty}^{\infty} \prod_{t=1}^{T} f(\epsilon_{it} \mid \alpha_i) \cdot f(\alpha_i) d\alpha_i \cdot d\epsilon_{iT} \cdot \cdots \cdot d\epsilon_{i1}$$

$$(52) \qquad = \int_{-\infty}^{\infty} f(\alpha_i) \prod_{t=1}^{T} [F(\infty|\alpha_i) - F(-x_{it}'\beta|\alpha_i)] \cdot d\alpha_i$$

$$(53) \qquad lnL = \sum lnL_i$$

$$(54) \qquad L_i = \int_{-\infty}^{\infty} f(\alpha_i) \prod_{t=1}^{T} F(x_{it}'\beta|\alpha_i) d\alpha_i$$

$$(55) \qquad L_i = \int_{-\infty}^{\infty} f(\alpha_i) \prod_{t=1}^{T} F(\tilde{y}_{it} - z_{it}) dv_i$$

wobei $\tilde{y}_{it} = 2y_{it} - 1; \quad v_i = u_i/\sqrt{2}; \quad z_{it} = [x_{it}'\beta + 2\rho v_i]/(1-\rho^2)^{1/2} =: x_{it}'\tilde{\beta} + \theta v_i =: \tilde{x}_{it}'\tilde{\tilde{\beta}}$

$$\partial L_i / \partial \tilde{\tilde{\beta}} = \int_{-\infty}^{\infty} f(\alpha_i) [\sum_{t=1}^{T} \tilde{y}_{it} \frac{f_{it}}{F_{it}}] \cdot (\prod_{t=1}^{T} F_{it}) \cdot \tilde{x}_{it} dv_i$$

wobei $\quad f_{it} = f(\tilde{y}_{it} z_{it}); \quad F_{it} = F(\tilde{y}_{it} z_{it}).$

Die Verbindung von Random-Effects-Modellen und Probit-Ansätzen ist naheliegend, da bei beiden Normalverteilung angenommen wird. Eine Alternative bietet der Minimum-Distanz-Schätzer

$$(56) \qquad (vec\hat{B} - f(\beta))'\hat{\Omega}(vec\hat{B} - f(\beta)) = Min.$$

wobei $\hat{B} = (\hat{\beta}_1, \ldots, \hat{\beta}_T)$.

Der Chamberlain-Ansatz geht von der logarithmierten Likelihood-Funktion aus

$$
\begin{aligned}
lnL &= \sum_{i=1}^{N} lnL_i \\
L_i &= L_i(y_{it}^*) = L_i(\epsilon_{it}; \alpha_i) \\
&= L_i(y_{it}^* | \alpha_i) \cdot L(\alpha_i) \\
&= \int_{-\infty}^{+\infty} \prod_{t=1}^{T} F(x_{it}'\beta + \alpha_i)^{y_{it}} \cdot (1 - F(x_{it}'\beta + \alpha_i))^{1-y_{it}} \cdot dG(\alpha_i),
\end{aligned}
$$

wobei $G(\alpha_i)$ eine eindimensionale Verteilungsfunktion ist, aus der α_i mit $i = 1, \cdots N$ stammt und $dG(\alpha_i)/d\alpha_i = f(\alpha_i)$, so dass $F(\alpha_i)d\alpha_i$ durch $dG(\alpha_i)$ ersetzt werden kann. Chamberlain (1984) geht von der Annahme

$$
\alpha_i = \sum_{t=1}^{T} a_t' x_{it} + \eta_i
$$

aus, so dass

$$
L_i = \int_{-\infty}^{\infty} \prod_{t=1}^{T} F(x_{it}'\beta + a'x_i + \eta_i)^{y_{it}} \cdot (1 - F(x_{it}'\beta + a'x_i + \eta_i)^{1-y_{it}} \cdot G^*(\eta_i)
$$

wobei $a = (a_1', \cdots, a_T')' = (a_{11}, \cdots, a_{K1}, \cdots, a_{1T}, \cdots, a_{KT})' \sim TK \times 1$, $F(\cdot)$ - Standardnormalverteilung und $G^*(\eta)$ - Normalverteilung mit $\eta \sim N(0, \sigma_\eta^2)$. Daraus folgt als multivariates Probitmodell mit

$$
y_{it} = 1, \text{wenn} \quad x_{it}'\beta + a'x_i + \eta_i + \epsilon_{it} > 0.
$$

Die Störgrößen seien unabhängig normalverteilt

$$
\begin{pmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{iT} \end{pmatrix} + \begin{pmatrix} \eta_i \\ \vdots \\ \eta_i \end{pmatrix} = \epsilon_i + \eta_i \iota \sim N\left[0, (I_T + \sigma_\eta^2 \iota\iota')\right],
$$

so dass sich aufgrund der multivariaten Normalverteilung von ϵ ein multivariates Probitmodell ergibt, bei dem ein T-dimensionales Integral zu lösen ist. Dies lässt sich vermeiden, wenn man die bezüglich x_i bedingte Verteilung von y_{it} bildet

$$
P(y_{it} = 1 | x_i) = F[(x_{it}'\beta + a'x_i)/(1 + \sigma_\eta^2)^{1/2}] = F(x_i'\pi),
$$

die aber als Randverteilung bezüglich α zu verstehen ist. Getrennt für jede Periode $t = 1, \cdots T$ lassen sich so univariate Probitspezifikationen als ML-Schätzung $(\hat{\Pi}_t = (\hat{\beta}, \hat{a}); t = 1, \cdots, T)$ bestimmen, die zusammengefasst gegen

$$
\Pi^* = diag\left[(1 + \sigma_\eta^2)^{-1/2}\right] (\underset{T \times TK}{I_T \otimes \beta'} \quad + \quad \underset{T \times 1}{\iota} \cdot \underset{1 \times TK}{a'}),
$$

konvergieren, wenn $N \to \infty$. Damit kann als Minimum-Distanz-Schätzung gebildet werden

$$(\hat{\pi} - f(\theta))'\hat{\Omega}^{-1}(\hat{\pi} - f(\theta)) = Min.,$$

wobei

$$\begin{array}{rcll} \underset{KT^2 \times 1}{\hat{\pi}} & = & vec(\hat{\Pi}) & \text{untereinandergesetzte, unrestringierte, univariate} \\ & & & \text{ML-Schätzung der T Probitmodelle} \\ \pi^* & = & vec(\Pi^*) = f(\theta) = f(\beta' \vdots a' \vdots \sigma_\eta^2)'. & \end{array}$$

$\hat{\Omega}$ ist ein konsistenter Schätzer der asymptotischen Kovarianzmatrix von $\hat{\Pi}$, d.h. von $\Omega = J^{-1}\Delta J^{-1}$

$$\underset{KT^2 \times T^2K}{J} = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_T \end{bmatrix} \quad ; \quad \underset{KT \times TK}{J_t} = E\left[\frac{f_{it}^2}{F_{it}(1-F_{it})} x_i x_i'\right]$$

J-Informationsmatrix, $\Delta = E[\psi_i \otimes x_i x_i']$; $\psi_i \sim (T \times T)$ Matrix mit den Elementen $\tilde{u}_{it} \cdot \tilde{u}_{it'}$, wobei $\tilde{u}_{it} = \frac{y_{it} - F_{it}}{F_{it}(1-F_{it})} f_{it}$ verallgemeinerte Residuen sind und $F_{it} = F(x_i'\pi), f_{it} = f(x_i'\pi)$. Als Schätzer für π_i ist $\hat{\pi}_i$ zu verstehen. Für Anwendungen sind die Erwartungswerte durch die entsprechenden arithmetischen Mittel und π durch $\hat{\pi}$ in f_{it} bzw. F_{it} zu ersetzen.

4.2.3 Zähldatenmodelle

Für Paneldatenmodelle mit Zähldaten $(0, 1, 2, \cdots, L)$ als endogene Variable wird am häufigsten ein Poissonansatz gewählt

$$P(y = j) = \frac{exp(-\lambda)\lambda^j}{j!} \qquad j = 0, 1, 2, ...$$

Abweichend vom linearen Modell erfolgt üblicherweise eine multiplikative Modellierung des Individualterms (Cameron/Trivedi 1998, S.275ff.). Das Grundmodell lautet dann

(57) $$y_{it} = \alpha_i \cdot \mu_{it} + u_{it}$$

mit

$$E[y_{it}|x_{it}, \alpha_i] = \lambda_{it} = \alpha_i \cdot exp(x_{it}'\beta) = exp(\tilde{\alpha}_i + x_{it}'\beta),$$

wobei $\tilde{\alpha}_i = ln\alpha_i$ und $i = 1, \cdots, N; t = 1, \cdots, T$. Unterstellt ist strikte Exogenität der Regressoren x. Im Falle fixer Individualeffekte kann wie bei entsprechenden Logit-Modellen von der CML-Methode Gebrauch gemacht werden, die

(58) $$lnL^c(\beta) = \sum_{i=1}^{N}[ln(\sum_{t=1}^{T} y_{it})! - \sum_{t=1}^{T} ln(y_{it}!) + \sum_{t=1}^{T} y_{it} ln[\frac{exp(x_{it}'\beta)}{\sum_{\tau=1}^{T} exp(x_{i\tau}'\beta)}]]$$

maximiert. Das c in $lnL^c(\beta)$ steht für "conditional". Als Bedingung geht ein, dass $\sum y_{it} = T\bar{y}_i$. Hausman, Hall und Griliches (1984) folgen diesem Vorgehen. Daneben schätzen sie auf gleichem Weg auch ein negatives Binomialverteilungsmodell mit fixen Effekten. Blundell, Griffith und Windmeijer (2002) zeigen, dass bei strikter Exogenität der Regressoren der CML-Schätzer für das Poissonmodell übereinstimmt mit dem Momentenschätzer des mittleren Skalenmodell

$$y_{it} = exp(x'_{it}\beta)\frac{\bar{y}_i}{\bar{\mu}_i} + u^*_{it},$$

wobei $\bar{y}_i = T^{-1}\sum_t y_{it}$, $\bar{\mu}_i = T^{-1}\sum_t \mu_{it}$ und $u^*_{it} = u_{it} - (\mu_{it}/\bar{\mu}_i)\bar{u}_i$. Das Verhältnis der beiden Mittelwerte misst den Individualeffekt. Der Momentenschätzer lautet

$$\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}u^*_{it} = \sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}(y_{it} - exp(x'_{it}\beta)\frac{\bar{y}_i}{\bar{\mu}_i}) = 0.$$

Bei nur schwacher und nicht strikter Exogenität, die gegenüber der schwachen Exogenität nicht zulässt, dass y_{t-1} grangerkausal zu x_t ist, ist die Momentenbedingung nicht mehr konsistent, da dann x_{it} und u^*_{it} über $\bar{u}_i$ korreliert sind. Einen Ausweg bietet die Quasi-Differenzenbildung der Momentenbedingung, so dass der Individualeffekt entfernt wird. Zu bilden sind folgende gewichtete Differenzen

$$w_{it} = y_{it} - \frac{\mu_{i,t-1}}{\mu_{i,t}} - y_{i,t-1} = u_{it}\frac{\mu_{i,t-1}}{\mu_{it}} - u_{i,t-1}.$$

Werden Instrumente $z_{i1}, \cdots, z_{it}$ verwendet, die die Bedingung

$$E[u_{it}|\alpha_i, z_{i1}, \cdots, z_{it}] = 0$$

erfüllen, so dass

$$E[w_{it}|z_{i1}, \cdots, z_{it}] = E_{\alpha|z}[E(w_{it}|z_{i1}, \cdots, z_{it})] = 0,$$

dann kann der GMM-Schätzer herangezog werden, der

$$(\frac{1}{N}\sum_{i=1}^{N} w'_i Z_i)W_N^{-1}(\frac{1}{N}\sum_{i=1}^{N} Z'_i w_i)$$

minimiert. Die Gewichtungsmatrix W_N entspricht

$$W_N = \frac{1}{N}\sum_{i=1}^{N} Z'_i w_i(\tilde{\beta})w_i(\tilde{\beta})'Z_i,$$

wobei $w_i(\tilde{\beta})$ ein konsistenter Anfangsschätzer ist. Bei exakt identifizierenden Instrumenten und $z_{it} = x_{it}$ entspricht die Stichprobenmomentenbedingung in der Struktur der Momentenbedingung bei strikter Exogenität

$$\sum_{i=1}^{N}\sum_{t=1}^{T-1} z_{i,t-1}w_{it} = \sum_{i=1}^{N}\sum_{t=1}^{T-1} z_{i,t-1}(y_{i,t-1} - exp(x'_{i,t-1}\beta)\frac{\bar{y}_{i,t-1}}{\bar{\mu}_{it}}) = 0.$$

Nur bei Regressoren mit größerer Variabilität kann dieser Schätzer befriedigen, wie Monte-Carlo-Studien zeigen. Schätzer mit Anfangswerten (Pre-Sample-Informationen) anstelle von

Zukunftswerten in der Gewichtung führen zu deutlichen Verbesserungen bei Simulationen, obwohl $\bar{u}_{it}^0$ immer noch mit x_{it} korreliert ist und somit der Schätzer verzerrt ist. Der Momentenschätzer, der die Stichprobenbedingung

$$\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-x_{it0})(y_{it}-exp(x_{it}'\beta)\frac{\bar{y}_{i0}}{\bar{\mu}_{i0}})=\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-x_{it0})u_{it}^0=0$$

löst, weist jedoch nur einen geringen Verzerrungsgrad auf. Weitere verteilungsfreie Schätzer für Paneldatenmodelle mit multiplikativem Individualterm, die auf Zähldatenmodelle, aber auch auf multinomiale Modelle anwendbar sind, stellt Wooldridge (1999) vor. Für Gaußsche Random-Effects-Poissonmodelle existieren keine analytischen Ausdrücke für die unbedingte Dichte. Ausweichen kann man hier auf Simulationsschätzer. Chip/Greenberg/Winkelmann (1998) nutzen daher z.B. die MCMC-Methode.

4.2.4 Selektionskorrekturmodelle

Methoden, die berücksichtigen, dass eine Beschränkung der Daten auf Teilgruppen zu verzerrten OLS-Schätzern führen, gehen auf Roy (1951) und Heckman (1974) zurück, sind von Lee (1982) verallgemeinert worden und wurden zuerst von Hausman/Wise (1979), Verbeek (1991), Verbeek und Nijman (1992) auf Random-Effects-Paneldatenmodelle übertragen. Fixed-Effects-Ansätze betrachten Verbeek (1991), Nijman/Verbeek (1992) und Zabel (1992). Eine systematische Diskussion zu Selektionskorrekturmodellen für Paneldaten bei fixen Effekten führt Wooldridge (1995). Ausgangspunkt bildet

(59) $$y_{it}=x_{it}'\beta+\alpha_i+u_{it}$$

als Basisgleichung und

(60) $$d_{it}^*=z_{it}'\gamma+\eta_i+\epsilon_{it};\quad d_{it}=I[d_{it}^*\geq 0]$$

als Selektionsgleichung, wobei $I[\cdot]$ eine Indikatorfunktion ist, die den Wert Eins annimmt, wenn das Argument in $[\cdot]$ erfüllt ist, und sonst Null. Beide unbeobachtbaren Individualterme α_i und η_i können mit den beobachtbaren Regressoren x_{it} und z_{it} korreliert sein. Als Spezialfall lässt sich $z_{it}=x_{it}$ formulieren. Der Regressand y_{it} ist nur beobachtbar, wenn $d_{it}=1$. Folgende Restriktionen sind auferlegt:

A1 : η_i ist linear abhängig von z_i:

$$\eta_i=z_{i1}\gamma_1+\cdots+z_{iT}\gamma_T+c_i$$

A2 : $\nu_{it}=\epsilon_{it}+c_i$ ist unabhängig von x_i und z_i; $\nu_{it}\sim N(0,\sigma_t^2)$.

A3 : α_i hängt linear ab von x_i und ν_{it}:

$$(\alpha_i|x_i,\nu_{it})=\psi_{t0}+x_{i1}'\psi_{t1}+\cdots+x_{iT}'\psi_{tT}+\phi_t\nu_{it}$$

A4 : u_{it} ist im Mittel unabhängig von x_i und z_i und sein bedingter Erwartungswert hängt linear von ν_{it} ab:

$$E(u_{it}|x_i,z_i,\nu_{it})=E(u_{it}|\nu_{it})=\rho_t\nu_{it}.$$

Zabel (1992) geht davon aus, dass α_i nur von x_i abhängt. Bei Gültigkeit der Annahmen A1-A4 lässt sich β aus

$$(61) \qquad E(y_{it}|x_i, z_i, \nu_{it}) = \psi_{t0} + x'_{i1}\psi_{t1} + \cdots + x'_{iT}\psi_{tT} + x'_{it}\beta + \delta_t\lambda(\frac{z'_i\gamma_t}{\sigma_{z'_i\gamma_t}})$$

mit $\delta_t = \rho_t + \phi_t$ identifizieren. Das Schätzverfahren läuft dann in folgenden Schritten ab

(i) Für jede Periode t ist die Selektionsgleichung mit Hilfe des Probitansatzes zu schätzen und daraus die Selektionskorrekturvariable $\hat{\lambda}(\cdot)$ zu bestimmen.

(ii) Der gepoolte OLS-Schätzer ist auf die um den Selektionsterm $\delta_t\lambda(\frac{z'_i\gamma_t}{\sigma_{z'_i\gamma}})$ erweiterte Basisregressionsgleichung

$$y_{it} = \hat{w}'_{it}\theta + \tilde{u}_{it}$$

mit $\hat{w}'_{it} = (1, x'_{i1}, ..., x'_{iT}, x_{it}, 0, ..., 0, \hat{\lambda}_{it}, 0, ..., 0)$ anzuwenden. Daraus ergibt sich

$$\hat{\theta} = (\hat{\psi}', \hat{\beta}', \hat{\delta}')' = (\sum_{i=1}^{N}\sum_{t=1}^{T} d_{it}\hat{w}_{it}\hat{w}'_{it})^{-1}(\sum_{i=1}^{N}\sum_{t=1}^{T} d_{it}\hat{w}_{it}y_{it}).$$

(iii) Die asymptotische Kovarianzmatrix für $\hat{\theta}$ ist vom Sandwich-Typ mit

$$V(\hat{\theta}) = \frac{1}{N}\hat{A}^{-1}\hat{B}\hat{A}^{-1},$$

wobei

$$\hat{A} = N^{-1}(\sum_{t=1}^{T} d_{it}\hat{w}'_{it}\hat{w}_{it}); \quad \hat{B} = N^{-1}(\hat{p}_i\hat{p}'_i); \quad \hat{p}_i = \hat{q}_i - \hat{D}\hat{r}_i; \quad \hat{q}_i = \sum_{t=1}^{T} d_{it}\hat{w}'_{it}\hat{e}_{it};$$

$$\hat{D} = N^{-1}(\sum_{t=1}^{T} d_{it}\hat{w}'_{it}\hat{\theta}'\nabla_{\delta}w_{it}(\hat{\delta}))'; \quad \hat{e} = y_{it} - \hat{w}'_{it}\hat{\theta}; \quad \hat{r}_i = r_i(\hat{\delta}),$$

wobei $\nabla_{\delta}w_{it}(\hat{\gamma}))'$ die Gradientenmatrix von $\hat{w}_{it}(\hat{\delta})'$ bewertet an der Stelle $\hat{\delta}$.

Einen gegenüber Wooldridge (1995) modifizierten Ansatz entwickelt Rochina-Barrachina (1997). Erstens werden Differenzen über alle Perioden gebildet, um den Individualterm zu beseitigen. Und zweitens wird eine trivariate Normalverteilung zwischen den Differenzen der Störgrößen $u_{it} - u_{i\tau}$ und den Störgrößen ν_{it}, $\nu_{i\tau}$ angenommen. Dies führt zu der Schätzgleichung

$$y_{it} - y_{i\tau} = (x_{it} - x_{i\tau})'\beta + \delta_t\lambda_{it}(\frac{z'_i\gamma_t}{\sigma_{z'_i\gamma_t}}, \frac{z'_i\gamma_\tau}{\sigma_{z'_i\gamma_\tau}}, \rho_{t\tau}) + \delta_\tau\lambda_{i\tau}(\frac{z'_i\gamma_\tau}{\sigma_{z'_i\gamma_\tau}}, \frac{z'_i\gamma_t}{\sigma_{z'_i\gamma_t}}, \rho_{\tau t}) + u_{it\tau}.$$

Um eine Schätzung für den Selektionskorrekturterm λ zu erhalten, werden aufgrund der reduzierten Form bivariate Probitschätzungen für jede Kombination von jeweils zwei Perioden durchgeführt. Im zweiten Schritt kann dann, wenn die λ-Werte durch ihre Schätzungen ersetzt worden sind, eine OLS-Schätzung für die Basisregressionsgleichung herangezogen werden. Ebenfalls mit der Annahme einer trivariaten Normalverteilung arbeitet Carrasco (2001). Allerdings ist keine Korrelation zwischen Regressoren und Individualeffekt zugelassen. Dies bedeutet, zufällige Effekte werden betrachtet. Außerdem ist die Basisgleichung kein linearer Ansatz, sondern ein Modell mit binären endogenen Variablen.

Vella und Verbeek (1999) entwickeln eine Conditional-Moment-Schätzung für ein Modell mit zensierter endogener Variablen in der Hauptgleichung und einer Selektionsgleichung, die gegenüber der Hauptgleichung zum Teil verzögerte Regressoren enthält. Für beide Gleichungen ist eine nichtlineare Modellierung der Regressoren zugelassen und es wird jeweils ein spezifischer individueller Random-Effect formuliert. Die unbeobachtete Heterogenität findet in die Hauptgleichung durch einen zusätzlichen geschätzten Regressor Eingang. Bei Vernachlässigung der verzögerten Regressoren lassen sich auch die Wooldridge-Ansätze verwenden.

Wenn davon ausgegangen wird, dass die Selektionsvariable d_{it} nicht endogen bestimmt wird, aber auch nicht zeitinvariant ist und gleiche Individualeffekte sowie gleiche Störgrößen bei zwei möglichen Regimen unterstellt werden

$$y_{it} = x_{it}'\beta_1 \cdot I[d_{it} \leq c] + x_{it}'\beta_2 \cdot I[d_{it} > c] + \alpha_i + u_{it}, \tag{62}$$

dann können die von Hansen (1999) vorgeschlagenen Kleinste-Quadrate-Schätzungen und die daraus resultierenden Testverfahren verwendet werden. Bei diesem Threshold-Modell, bei dem der Regimewechsel bei dem konstanten Wert c eintritt, wird der Individualeffekt durch eine Within-Modellierung beseitigt. Bei gegebenem c ist die OLS-Schätzung von β geeignet. Ansonsten wird

$$\hat{c} = argmin[(\hat{u}_{it}(c) - \bar{\hat{u}}_i(c))'(\hat{u}_{it}(c) - \bar{\hat{u}}_i(c))] \tag{63}$$

empfohlen, wobei $\hat{u}_{it}$ den OLS-Residuen bei alternativen c-Werten entspricht, ein Vorgehen, das z.B. ähnlich bei Autokorrelationskoeffizienten von Hildreth und Lu (1960) vorgeschlagen wurde. Hansen erweitert seinen Ansatz auf mehr als zwei Regime.

4.2.5 Empirische Studien

Auf den schon klassisch zu nennenden Anwendungsfall für Zähldatenmodelle, Zahl der Patente, gehen Blundell, Griffith und Windmeijer (2002) ein. Den klassischen Fall für Selektionsmodelle, den Zusammenhang zwischen Arbeitsangebot und Löhnen, wählen Vella und Verbeek (1999) als Anwendungsbeispiel. Gleiches gilt für Dustmann/Rochina-Barrachina (2000). Carrasco (2001) untersucht den Einfluss der Fertilität auf die Erwerbsbeteiligung von Frauen. Investitionen und Finanzierungsbeschränkungen unter Verwendung US-amerikanischer Paneldaten für 565 Firmen über 15 Jahre kommen bei Hansen (1999) im Rahmen von Regimewechselmodellen zur Anwendung.

5 Nicht- und semiparametrische Modelle

Nichtparametrische Ansätze haben ihren Ausgangspunkt bei Histogrammen. Kernschätzer und andere Verfahren stellen eine verallgemeinerte Methode zur Bestimmung von Dichtefunktionen dar und lassen sich auf Regressionsmodelle anwenden. Während bei parametrischen Modellen von einer bekannten strukturellen Beziehung unter Berücksichtigung der unbekannten Parameter und der Störgröße ausgegangen wird ($y = g(x, \beta, u)$), formulieren nichtparametrische Regressionsmodelle lediglich eine Mittelwertregression zwischen den Variablen y und x

$$y = E(y|x) + u = m(x) + u, \tag{64}$$

wobei x die unabhängige Variable ist, die in einer unbekannten funktionalen Beziehung zu y steht $m(x) = \int y f(y|x) dy$, und u den Störterm mit E(u|x)=0 beschreibt. Mit Hilfe nichtparametrischer Verfahren kann $m(x)$ ohne weitere Informationen punktweise geschätzt werden. Es handelt sich um einen lokal linearen Regressionsansatz. Die Motivation, nichtparametrische statt parametrische Methoden zu verwenden, besteht in der größeren Flexibilität. Es werden keine Verteilungsannahmen auferlegt. Die ökonomische Theorie kann üblicherweise keine Aussagen über die Verteilung machen. Auch ist a priori nur in wenigen Fällen der Zusammenhang zwischen zwei Variablen genau vorgegeben. Wird aber trotz allem eine exakt spezifizierte Beziehung bei der Schätzung zugrunde gelegt, so ist die Gefahr von Fehlspezifikation erheblich. Der Vorteil der nichtparametrischen Modellierung besteht darin, dass sich unter relativ allgemeinen Bedingungen konsistente Schätzungen erzielen lassen. Ein Nachteil parametrischer Schätzungen ist, dass statistisch gesicherte Abweichungen von der unterstellten Verteilung der Störgrößen keinen Hinweis dafür liefern, welcher Verteilungsfamilie die Störgrößen tatsächlich angehören. Wenn jedoch die funktionale Beziehung bekannt ist, dann ergibt sich gegenüber parametrischen Ansätzen ein Effizienzverlust. Nichtparametrische Ansätze sind eher statistisch-explorativ als ökonomisch-theoretisch orientiert. Nichtparametrische Regressionsschätzungen können als Instrument zur Aufdeckung von Nichtlinearitäten dienen. In der empirischen Wirtschaftsforschung haben sich nichtparametrische Verfahren noch nicht so recht durchgesetzt. Von Ausnahmen abgesehen fehlt es an der Implementation dieser Verfahren in ökonometrischen Softwareprogrammen. Nachteilig an dieser Vorgehensweise ist der hohe Rechenaufwand. Dieses Problem verliert durch leistungsfähige Rechner zunehmend an Bedeutung. Der Aufwand nimmt mit der Zahl der Beobachtungen überproportional zu. Große Stichprobenumfänge werden aber benötigt, um hinreichend gute Glättungen zu erzielen. Die Konvergenzrate liegt häufig sehr niedrig. An einer allgemein akzeptierten Methode zur Bestimmung der Bandweite, die erheblichen Einfluss auf die Glättung hat, fehlt es bisher. Weitere Probleme ergeben sich bei nichtparametrischen Schätzern, wenn die Zahl der Regressoren zunimmt. In diesem Zusammenhang wird vom Fluch der Dimensionalität gesprochen. Jeder zusätzliche Regressor erhöht den Umfang der Berechnungen erheblich und die Schätzer werden unpräzise. Es werden überdimensional große Stichproben benötigt, um eine akzeptable Schätzpräzision zu erhalten. Schwierigkeiten bei nichtparametrischen Schätzern ergeben sich mit zunehmender Zahl der Regressoren auch bei der graphischen Darstellung und der Interpretation. Schließlich sind auch praktisch keine Möglichkeiten der Prognose gegeben. Dem Trade-off zwischen Flexibilität und Handhabbarkeit der Regressionsbeziehung lässt sich durch zwei alternative Vorgehensweisen begegnen. Entweder es wird ein **additives Modell** nichtparametrischer Terme gebildet

$$y = \sum_{k=1}^{K} m_k(x_k) + u, \tag{65}$$

oder der nichtparametrische Teil konzentriert sich auf eine Schlüsselvariable, während der Rest parametrisch, üblicherweise linear modelliert wird

$$y = m(x) + z'\gamma + u. \tag{66}$$

Dieses Modell heißt **partiell linear** oder semilinear. Der Begriff **semiparametrisch** wird in der Literatur nicht einheitlich gefasst. Die weitestgehende Definition beschreibt semiparametrische Modelle einfach als solche, die parametrische und nichtparametrische Komponenten enthalten. Die Anwendung semiparametrischer Ansätze bietet sich außer bei partiell linearen Modellen dann an, wenn die Verteilung der Störgrößen unbekannt ist, wenn Linearkombinationen von Einflussfaktoren in unbekannter, nichtlinearer Form auf den Regressanden wirken, wenn die endogene Variable eine latente Variable ist oder wenn das Modell generierte Regressoren enthält wie bedingte Erwartungswerte, die unbekannt sind. Einen Überblick zu nicht- und semiparametrischen sowie zu additiven Modellen liefern Buja/Hastie/Tibshirani (1989), Härdle (1990), Hasti/Tibshirani (1990), Ullah/Vinod (1993), Lee(1996), Horowitz (1998), Pagan/Ullah (1999), Härdle/Liang/Gao (2000), König (2002).

5.1 Ansätze für Querschnittsdaten

5.1.1 Nichtparametrische Modelle

Die ersten nicht- und semiparametrischen Schätzungen haben sich auf Querschnittsdatenmodelle konzentriert. Die Verfahren sind verstreut in der Literatur zu finden und weisen einen erheblichen Heterogenitätsgrad auf. Die Schätzer sind teilweise auf ganz spezielle Modelle zugeschnitten und zielen teilweise auf einen sehr hohen Allgemeinheitsgrad ab. Sehr häufig wird bei den nichtparametrischen Schätzern die Glättungsfunktion über ein lokal gewichtetes arithmetisches Mittel bestimmt

$$\hat{m}(x) = \sum_{i=1}^{N} w_i(x) y_i. \tag{67}$$

In allgemeiner Formulierung des Kerndichteschätzers für die Dichtefunktion $f(x)$ eines K- dimensionalen Vektors lässt sich der nichtparametrische Term durch

$$\hat{f}(x) = \frac{1}{N} \sum_{i=1}^{N} K_H(x_i - x) \tag{68}$$

darstellen, wobei $K_H(x) = det(H)^{-1} K(H^{-1}x)$ mit einer multivariaten Kernfunktion K(x) bei gegebener Bandweitenmatrix H. Der multivariate Kern vereinfacht sich erheblich, wenn er sich als Produkt univariater Kerne der Form $K_H = \Pi_{k=1}^{K} K(x_k)$ darstellen lässt und die bandweitenmatrix eine Diagonalmatrix ist. Im univariaten Fall ist H ein Bandweitenskalar h und $\hat{f}_h(x)$ der univariate Kerndichteschätzer, so dass

$$\hat{f}(x) = \frac{1}{N \cdot h} \sum_{i=1}^{N} K(\frac{x_i - x}{h}).$$

Die Wahl des Kerndichteschätzers (z.B. $K(\frac{x_i - x}{h}) =: K(u_i) = \frac{1}{\sqrt{2\pi}} exp(-\frac{1}{2}u_i^2))$ als Gaußkern oder $K(u_i) = \frac{3}{4}(1 - u_i^2) I(u_i| \leq 1)$ als Epanechnikov-Kern) ist weniger entscheidend als die der Bandweite h für das Schätzergebnis. Vorgeschlagen wird, h so festzulegen, dass der mittlere integrierte quadratische Fehler

$$MISE(\hat{f}(x)) = \int E[(\hat{f}(x) - f(x))^2] \cdot \pi(x) dx$$

minimiert wird, wobei f(x) die wahre Dichtefunktion und $\pi(x)$ eine Gewichtungsfunktion darstellt, die Beobachtungen am Rand eine geringere Bedeutung zuweist. In der Praxis wird üblicherweise das Stichprobenpendant, der mittlere quadratische Fehler $ASE = 1/N \sum(\hat{f}(x) - f(x))^2 \cdot \pi(x)$ verwendet. Falls die wahre Dichte normalverteilt ist, ergibt sich im univariaten Fall als optimaler Wert $1,06\sigma_x N^{-1/5}$.

Im Regressionsmodell führt für $\hat{m}(x)$ der bekannte Nadaraya-Watson-Schätzer zu

$$\hat{m}(x) = \sum_{i=1}^{N} \frac{K_h(x_i - x)}{\sum_{j=1}^{N} K_h(x_j - x)} \cdot y_i. \tag{69}$$

Gourieroux, Monfort und Tenreiro (2000) entwickeln einen allgemeinen nichtparametrischen Schätzer, der verschiedene bekannte Spezialfälle wie den Nadaraya-Watson-Kernschätzer, lokal lineare und lokal polynomiale Schätzer enthält und als Kern-M-Schätzer bezeichnet wird.

Ob nichtparametrische Verfahren parametrischen vorzuziehen sind, lässt sich mit Hilfe verschiedener Tests überprüfen. Ait-Sahilia, Bickel und Stoker (1994) entwickeln eine asymptotisch normalverteilte Teststatistik, die in der Struktur dem ASE entspricht, wobei zwei nichtparametrische Schätzungen miteinander verglichen werden, d.h. eines restringierten mit einem weniger restringierten Modell. Als Spezialfall ist darin auch der Vergleich eines parametrischen mit einem nichtparametrischen Schätzer enthalten. Liegt ein lineares Modell vor, ist der parametrische Schätzer $x'\hat{\beta}$ mit $\hat{m}(x)$ zu vergleichen. Einen alternativen Test auf die funktionale Form eines nichtlinearen Regressionsmodells mit Hilfe eines kombinierten CM-Tests und eines nichtparametrischen Tests schlägt Zheng (1997) vor. Fehlspezifikationstests, die sich am parametrischen Vorgehen orientieren, präsentieren Ullah und Vinod (1993, 106ff). Die Frage zur Diskriminierung zweier nicht genesteter Erklärungsansätze (Variablensätze) beantworten Lavergne und Vuong (1996), ohne dass die funktionale Form der Regression noch die Verteilung der Residuen spezifiziert ist. Ihre Teststatistik enthält den empirischen MSE aufgrund nichtparametrischer Kernschätzer. Die Residuenvarianz wird konsistent durch den empirischen MSE geschätzt. Die Teststatistik ist asymptotisch normalverteilt und im semiparametrischen Sinne asymptotisch effizient.

5.1.2 Semiparametrische Modelle

Bei **partiell linearen** Ansätzen der Gestalt (70) wird zunächst aufgrund der Identität E[m(x)|x]= m(x) über Differenzenbildung - Robinson (1988) folgend -

$$y - E(y|x) = (z - E(z|x))'\gamma + u \tag{70}$$

vom nichtparametrischen Term bereinigt. Naheliegend ist, eine einfache KQ-Schätzung dieses Ansatzes, wobei die unbekannten bedingten Erwartungswerte durch Schätzwerte $\hat{y}$ und $\hat{z}$ ersetzt werden. Es bietet sich an, diese nichtparametrisch konsistent zu schätzen (Pagan/Ullah 1999,199). Das Problem hierbei ist, dass die Nenner der Kernregressionsschätzungen Zufallsvariable sind:

$$\hat{y}_i = \hat{E}(y_i|x_i) = \frac{1}{N \cdot h^K} \sum_{j=1}^{N} \frac{K\frac{x_i - x_j}{h}}{\hat{f}_i} \cdot y_j,$$

wobei $\hat{f}_i = \frac{1}{N \cdot h^K} \sum_{j=1}^{N} K(\frac{x_i - x_j}{h})$ die Zufallsvariable ist. Für den multivariaten Kern wird die vereinfachte Form des Produktes der univariaten Kerne zugrunde gelegt, d.h. $K(x_i) = \Pi_{k=1}^{K} K(x_{ik})$. Entsprechendes gilt bei der Betrachtung des geschätzten bedingten Erwartungswertvektors von z_i, d.h. bei $\hat{z}_i$. Der Nenner lässt sich bei einer dichtegewichteten Schätzung (Powell/Stock/Stoker 1998, Li/Stengor 1996) vermeiden, wobei die Dichtefunktion f durch eine Kerndichteschätzung bestimmt wird ($\hat{f}$). Mit dieser ist das Differenzenmodell (70) zu gewichten

$$\hat{f}_i(y_i - \hat{y}_i) = \hat{f}_i(z_i - \hat{z}_i) + \hat{f}_i u_i,$$

um dann den KQ-Schätzer für γ zu bilden

$$\hat{\gamma} = (\sum_{i=1}^{N}(z_i - \hat{z}_i)(z_i - \hat{z}_i)'\hat{f}_i^2)^{-1}(\sum_{i=1}^{N}(z_i - \hat{z}_i)(y_i - \hat{y}_i)\hat{f}_i^2). \tag{71}$$

In einem zweiten Schritt kann dann der nichtparametrische Term $m(x)$ über eine Taylorreihenapproximation

$$y_i - z_i'\hat{\gamma} = m(x) + \beta(x)'(x_i - x) + R(x_i, x) + z_i'(\gamma - \hat{\gamma}) + u_i \tag{72}$$

mit Hilfe eines lokalen KQ-Schätzers bestimmt werden, d.h. $m(x)$ und $\beta(x)$ werden auf diesem Weg bestimmt. Eine Verbesserung lässt sich gegenüber dem dichtefunktionsgewichteten semiparametrischen Schätzer mit Hilfe eines Projection-Pursuit-Regressionsschätzers erzielen, wenn eine höhere Dimensionalität von x vorliegt, wie Monte-Carlo-Studien zeigen (Li/Stengos 2002). You und Chen (2002) betrachten die Parameterschätzung im partiell linearen Regressionsmodell für den Fall autoregressiver Störgrößen Autoregressionskoeffizienten, die zufallsbedingten Schwankungen unterliegen.

Spezielle semiparametrische Schätzer sind für **binäre Responsemodelle** der Gestalt

$$y^* = x'\beta + u \tag{73}$$

$$y = \begin{cases} 1, & wenn \quad y^* > 0 \\ 0 & sonst \end{cases}$$

entwickelt worden. Der Wert einer solchen Schätzung liegt bei diesem Modelltyp vor allem darin, dass bei Verletzung der zugrunde gelegten Verteilungsannahme parametrische Schätzer inkonsistent sind. Der bekannteste Ansatz geht auf Klein und Spady (1993) zurück, die eine Single-Index-Spezifikation wählen. Die Linearkombination im traditionellen Probit- oder Logitmodell $x'\beta$ wird als eine Variable aufgefasst, wobei $P(y = 1) = M(x'\beta)$ eine Funktion ist, die im Bereich [0,1] liegt. Der Koeffizientenvektor β kann dann, wenn $m(x)$ durch die nichtparametrische Schätzung $\hat{m}(x)$ ersetzt wird, über die Maximierung der (Quasi-)Log-Likelihoodfunktion ermittelt werden

$$lnL(\beta) = \frac{1}{N}\sum_{i=1}^{N}[w_i\{y_i \cdot lnM(x_i'\beta) + (1 - y_i) \cdot ln(1 - M(x_i'\beta)\}], \tag{74}$$

wobei y eine binäre [0;1]-Variable ist und w_i wie oben - (71) - einer Gewichtungsfunktion entspricht, die weiter entfernte Werte geringer gewichtet. Der Term $M(x_i'\beta)$ wird durch $\hat{M}(x_i'\beta)$ ersetzt und als gewichteter Anteilswert ermittelt aus

$$\hat{M}(x'\beta) = \frac{p_N \cdot \hat{m}_N(x'\beta|y = 1)}{p_N \cdot \hat{m}_N(x'\beta|y = 1) + (1 - p_N) \cdot \hat{m}_N(x'\beta|y = 0)}.$$

Bei N Beobachtungen entspricht p_N dem Anteilswert der Elemente mit $y = 1$ in der Stichprobe. Es ist offensichtlich, dass übliche Probit- und Logit-ML-Schätzer Spezialfälle des Klein-Spady-Ansatzes darstellen. Alternative Verfahren sind von Horowitz (1992), Lewbel (1997) und Matzkin (1993) entwickelt worden.

Lee(1999) entwickelt ein semiparametrisches bivariates binäres Response-Modell ohne Spezifizierung der Verteilung. Zur Bestimmung der asymptotischen Verteilung wird ein nichtparametrischer Schätzer und ein Test, ob zwischen den beiden diskreten Variablen Unabhängigkeit besteht, vorgeschlagen.

Für den Modelltyp mit **generierten Regressoren** formuliert Ahn (1997) ein Single-Index-Modell. Es wird vermutet, dass der bedingte Ewartungswert der Variablen y, die von dem nichtparametrischen Variablenvektor x und dem linear parametrischen Vektor z abhängt, sich durch den gewichteten Index

$$E(y|x,z) = E(y|x) \cdot \delta_1 + (z'\gamma) \cdot \delta_2 + =: v'\alpha$$

beschreiben lässt. Der nichtparametrische bedingte Erwartungswert $E(y|x) = m(x)$ sei unbekannt. Daher ist $E(v'\alpha) = g(v'\alpha)$ nur partiell beobachtbar. Die Annahme über den Index $v'\alpha$ impliziert

$$E(y|x,z)) = E(y|v'\alpha) = E(y|v),$$

da α ein Vektor von Konstanten ist. Dieser Zusammenhang gilt auch in einer um eine Gewichtungsfunktion $w(v)$ erweiterten Beziehung. Wird $E(y|v)$ allgemein durch $\tilde{g}(v)$ notiert, so ergibt sich als Average-Derivative-Ansatz

$$E[w(v) \cdot \frac{\partial \tilde{g}(v)}{\partial v}] \cdot \alpha =: \gamma_w \cdot \alpha =: \delta$$

mit $w(v)$ als beliebige Gewichtsfunktion und $\gamma_w = E(w(v)g'(v'\alpha))$. Die Schätzung verläuft zweistufig. Im ersten Schritt wird der bedingte Erwartungswert $E(y|x)$ nichtparametrisch mit Hilfe eines Kernschätzers bestimmt ($E(y|x) = m(x)$). Dann ist v_i durch $\hat{v}_i = (\hat{m}_i, z_i)$ zu ersetzen. Der zweite Schritt besteht in der Schätzung von δ, d.h. des bis auf einen skalaren Faktor identifizierbaren Parametervektors α. Gefolgt wird dabei einem Ansatz von Powell/Stock/Stoker (1989)

$$\hat{\delta} = -2\frac{1}{N}\sum_{i=1}^{N}\frac{\partial \hat{f}(v_i)}{\partial v_i} \cdot y_i, \tag{75}$$

wobei $f(v)$ der gewählten Gewichtungsfunktion $w(v)$ entspricht. Zur Bestimmung von $\hat{f}(v)$ eignet sich wiederum ein Kernschätzer.

Semiparametrische Modelle zur Berücksichtigung des **Sample-Selection-Bias** sind motiviert durch die Kritik an den parametrischen Ansätzen dieses Typs. Die künstliche Selektionsvariable und die eigentlichen Regressoren sind häufig hoch kollinear, insbesondere im zentralen Wertebereich, während an den Rändern nichtlineare Zusammenhänge zum Ausdruck kommen. Aus einem hohen Multikollinearitätsgrad resultieren instabile, nichtrobuste und stark sensitiv auf kleine Änderungen reagierende Schätzungen. Wie Monte-Carlo-Studien zeigen, werden sehr große Standardfehler für die Schätzfunktionen der Koeffizienten erzeugt. Noch wichtiger ist, dass die Ergebnisse stark von den Verteilungsannahmen beeinflusst werden. Naheliegend ist, von dem üblichen Selektionsmodell

$$y^* = z'\gamma + u$$

$$d^* = x'\beta + \epsilon$$

$$d = \begin{cases} 1, & wenn \quad d^* > 0 \\ 0 & sonst \end{cases}$$

$$y = y^* \cdot d$$

abzugehen, das eine Standardnormalverteilung der Störgrößen unterstellt und daher zu dem Ergebnis

(76) $$E[u|x, d = 1] = \sigma_{u\epsilon} \frac{\phi(x'\beta)}{1 - \Phi(x'\beta)}$$

kommt. Um eine Verteilungsannahme zu vermeiden, wird stattdessen von

(77) $$E[u|x, d = 1] = m(x'\beta)$$

ausgegangen, wobei $m(x'\beta)$ eine unbekannte Funktion ist. Dies entspricht einer Single-Index-Formulierung. Die Probleme gegenüber der parametrischen Schätzung sind offensichtlich. Erstens lässt sich β nicht mehr unter Berücksichtigung der Normalverteilungsannahme von ϵ über ein Probitmodell schätzen. Und zweitens kann damit auch nicht mehr der Selektionsterm $E[u|x, d = 1]$ ermittelt werden. Das erste Problem kann mit Hilfe der voranstehenden nichtparametrischen Methode für Discrete-Choice-Modelle gelöst werden. Für das zweite Problem bieten sich verschiedene Lösungsmöglichkeiten an (Vella 1998, 139ff). Auszugehen ist von

(78) $$E[y|x, d = 1] = z'\gamma + m(x'\hat{\beta}).$$

Ein Lösungsverfahren ist oben für partiell lineare Modelle beschrieben.

Auch für **zensierte Regressionsmodelle**

(79) $$y^* = x'\beta + u$$

$$y = \begin{cases} y^*, & wenn \quad y^* > 0 \\ 0 & sonst \end{cases}$$

ist eine Reihe semiparametrischer Ansätze entwickelt worden. Insbesondere drei Typen von Schätzern sind zu unterscheiden:

- Schätzung nach dem Prinzip der Minimierung zensierter absoluter Abweichungen (CLAD)
- symmetrisch zensierter Kleinste-Quadrate-Schätzer (SCLS)
- paarweiser Differenzen-Schätzer für zensierte Beobachtungen (CPD).

Einen Überblick zu diesen Vorgehensweisen liefern Chay/Honore (1998) und Chay/Powell (2001). Der CLAD-Schätzer ist durch

$$\hat{\beta}_{CLAD} = argmin \sum_{i=1}^{N} |y_i - max(0, x'\hat{\beta})| \tag{80}$$

definiert. Der SCLS-Schätzer wird über

$$\hat{\beta}_{SCL} = argmin \sum_{i=1}^{N} (y_i - max\{\frac{1}{2}y_i, x_i'\hat{\beta}\})^2 + I[y_i > 2x_i'\hat{\beta}](\frac{1}{4}y_i^2 - max\{0, x_i'\hat{\beta}\}^2) \tag{81}$$

formuliert. Die Symmetrieeigenschaft ist bei zensierten Beobachtungen nicht notwendigerweise gegeben. Damit der SCLS konsistente Schätzungen liefert, ist davon auszugehen, dass die Störgröße einen Median von Null hat. Die Zensierung hat zur Folge, falls zwei Beobachtungen (i und j) betrachtet werden, dass $y_i - x_i'\beta = max\{-x_i'\beta, u_i\}$ und $y_j - x_j'\beta = max\{-x_j'\beta, u_j\}$ nicht zwangsläufig identisch verteilt sind, selbst wenn dies für u_i und u_j gilt. Wenn aber künstlich zensiert wird, indem anstelle von $max\{-x_i'\beta, u_i\}$

$$max\{-x_j'\beta, -x_i'\beta, u_i\}$$

gebildet wird sowie der analoge Ausdruck für die Beobachtung j, dann sind $y_i - x_i'\beta$ und $y_j - x_j'\beta$ unabhängig und identisch verteilt, vorausgesetzt u_i und u_j sind identisch, unabhängig verteilt. Unter dieser Bedingung für x_i und x_j folgt, dass

$$(max\{y_i, x_i'\beta - x_j'\beta\} - x_i'\beta) - (max\{y_j, x_j'\beta - x_i'\beta\} - x_j'\beta) =$$

$$max\{-x_i'\beta, -x_j'\beta, u_i\} - max\{-x_i'\beta, -x_j'\beta, u_j\}$$

symmetrisch um den Wert Null verteilt ist. Daher gilt

$$E[\xi(max\{y_i, (x_i - x_j)'\beta\} - max\{y_j, (x_j - x_i)'\beta\} - (x_i - x_j)'\beta)(x_i - x_j)] = 0,$$

wobei ξ irgendeine nichtsymmetrische Funktion ist, für die der Erwartungswert existiert. Diese Momentenbedingung kann genutzt werden, um den CPD-Schätzer zu definieren. Allerdings ist es für Anwendungszwecke notwendig, einen Schätzer auf Basis des analogen Stichprobenzusammenhangs zu formulieren

$$\hat{\beta}_{CPD} = argmin \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} s(y_i, y_j, (x_i - x_j)'\hat{\beta}), \tag{82}$$

wobei $s(\cdot)$ für drei Bereiche definiert ist.

Für ein erweitertes Tobitmodell, d.h. für ein **Zweigleichungsmodell mit zensierten Beobachtungen**,

$$y^* = x'\beta + u_1 \tag{83}$$

$$w^* = z'\gamma + u_2$$

$$(y,w) = \begin{cases} (y^*, w^*), & wenn\ y^* > 0 \\ (0,0) & sonst \end{cases}$$

entwickeln Honore, Kyriazidou und Udry (1997), Kyriazidou (1995) einen zweistufigen Schätzer, die vom CPD-Schätzer Gebrauch machen. Einen alternativen semiparametrischen Schätzer für das gleiche Modell, bei dem ebenfalls zweistufig vorgegangen wird, präsentiert Chen (1997). Semiparametrische Schätzer für **Zähldatenmodelle** finden sich bei Gurmu, Rilstone und Stern (1999). Für nichtlineare **gemischte Modelle**, die Elemente mit fixen und zufallsbedingten variablen Koeffizienten enthalten, zeigen Ke und Wang (2001) Möglichkeiten der Schätzung auf. Einen Test, ob eine semiparametrische gegenüber einer nichtparametrischen Modellierung vorzuziehen ist, findet sich bei Fan und Li (1996). Li und Wang (1998) bieten dagegen einen Ansatz an, bei dem semiparametrische Spezifikationen gegen lineare Modelle getestet werden.

5.1.3 Additive Modelle

Gegenüber multiplen linearen Regressionsmodellen stellen additive Modelle eine Verallgemeinerung dar. Einige oder alle linearen Regressoren zur Erklärung von y werden ersetzt durch beliebige Glättungsfunktionen, wobei die exogenen Variablen weiterhin additiv miteinander verknüpft sind, d. h. $y = X\beta + u$ wird ersetzt durch

(84) $$y = \beta_0 + \sum_{k=1}^{K} f_k(x_k) + u.$$

Bei der allgemeineren Modellierung

(85) $$y = f(x_1, \ldots, x_K) + u$$

taucht bei der Suche nach einer Lösung das Problem des Dimensionalität auf: Im mehrdimensionalen Raum ist zu entscheiden, welche bei verschiedenen Punktekombinationen die größte Nähe (Nachbarschaft) zu einem Bezugspunkt aufweist. Eine Vereinfachung ergibt sich zwar, wenn eine sogenannte "projection-pursuit regression", ein additives Modell von Prädiktoren

(86) $$y = \sum_{p=1}^{P} f_p(x'\alpha_p) + u$$

gebildet wird, bei dem der Verlauf verschiedener Projektionssterme verfolgt wird. In diesem Fall ist $x'\alpha_p$ jeweils eine eindimensionale Projektion des Vektors x. Den Spezialfall des einfachen additiven Modells erhält man, wenn die Richtungsvektoren α_p Einheitsvektoren sind und P=K entspricht. Im allgemeinen Fall sind Interaktionen zwischen den Regressoren zugelassen. Mit Hilfe einer Optimierungstechnik sind "gute" Richtungsvektoren $\alpha_1, \alpha_2, ..., \alpha_P$ zu finden. P wird vorgegeben und der mittlere quadratische Fehler

$$E[y - \sum_{p=1}^{P} f_p(x'\alpha_p)]^2$$

ist zu minimieren in Bezug auf α_p und f_p $\forall$ p. Üblicherweise entsteht ein Interpretationsproblem, da für jeden Summanden p eine Linearkombination mit den gleichen Regressoren die Grundlage für die Projektion auf y bildet. Wie lässt sich dann p=1 von p=2 usw. inhaltlich trennen?

In additiven Modellen ist das Problem gelöst. Hier wird getrennt für jeden Regressor eine nichtlineare Beziehung mit y hergestellt. Jede Funktion wird geschätzt durch Glättung an einer einzigen Koordinate. In diesem Fall können hinreichend viele Punkte in der Nachbarschaft eines Bezugspunktes x_0 einbezogen werden, um die Varianz niedrig halten zu können, auch wenn sie je nach Bezugspunkt x_0 sehr unterschiedlich sein kann. Je mehr Punkte in der Nachbarschaft von x_0 liegen, um so geringer ist die Varianz zu erwarten, um so besser wird der (nichtlineare) Verlauf, die Projektion von x auf y, bestimmt.

Additive Modelle können zwar zu verzerrten Schätzern für die wahre Regressionsoberfläche führen, es wird aber vermutet, dass der Bias geringer als bei einer hochdimensionalen Glättung ist (Hastie/Tibshirani 1986, 305). Modelle dieser Art stellen also einerseits eine Verallgemeinerung der linearen Modelle dar und lösen andererseits den "Fluch der Dimensionalität" auf, lassen einfache Interpretationen zu, welchen Beitrag die einzelnen Regressoren leisten. In der Praxis wird häufig eine Mischung aus linearem und allgemein additivem Modell sinnvoll sein

$$y = \beta_0 + \sum_{k=1}^{K_1} f(x_k) + \sum_{k=K_1+1}^{K_1+K_2} x_k\beta_k + u. \tag{87}$$

Die Schätzung additiver Regressionsmodelle kann auf vielen Wegen erfolgen. Die Unterschiede bestehen in der Art der Glättung, in welcher Form Glättungsrestriktionen den Funktionen auferlegt werden (vgl. Hastie/Tibshirani 1997, 89f.).

Der Backfitting-Algorithm ist ein allgemeines Lösungsverfahren, das es ermöglicht, eine additives Modell an die Daten anzupassen. Das Vorgehen ist iterativ. Angenommen

$$y = \beta_0 + \sum_{k=1}^{K} f_k(x_k) + u$$

ist korrekt und $f_1, ..., f_{j-1}, f_{j+1}, ..., f_k$ sind bekannt, dann lässt sich

$$u_j = y - \beta_0 - \sum_{k \neq j} f_k(x_k) \tag{88}$$

als partielles Residuum definieren. Dem Kleinste-Quadrate-Prinzip folgend, wird

$$E(y - \beta_0 - \sum_{k=1}^{K} f_k(x_k))^2 \tag{89}$$

minimiert, um f_j zu schätzen. Der Algorithmus geht von der Anfangsannahme $f_1^{(1)} = f_2^{(1)} = ... = f_K^{(1)} = 0$ aus. Diese führt zu *Stufe 1* des Verfahrens

$$\beta_0^1 = E(y).$$

In *Stufe 2* wird neu definiert

$$u_j^{(2)} = y - \beta_0 - \sum_{k=1}^{j-1} f_k^{(2)}(x_k) - \sum_{k=j+1}^{K} f_k^{(1)}(x_k).$$

Dieser Ausdruck ist für $j = 1, ..., K$ zu bestimmen, wobei $f_k^{(2)} = E(u_j|x_j)$. Die Schätzung von $f_k^{(2)}$ kann mit Hilfe verschiedener Verfahren erfolgen, z.B. durch Kernschätzer, lokal gewichtete Regressionsglättung, Splineregressionsglättung als nichtparametrische Verfahren oder über parametrische Ansätze wie Polynome und kubische Splines.

Wird das lokal gewichtete Scatter-Plot-Glättungsverfahren gewählt, so ist für $f_k^{(2)} =: f(x)$ und $u_j^{(2)} = y^*$ wie folgt zu verfahren, wobei $y^* = f(x) + u$ den Ausgangspunkt bildet:

(i) Wähle einen Beobachtungspunkt x_0 aus $x_1, \cdots, x_N$ und suche die n_0 nächsten Nachbarn von x_0, wodurch eine Nachbarschaft $N(x_0)$ definiert ist. Die Zahl der Nachbarn n_0 ist festgelegt als Prozentsatz aller Beobachtungen (span).

(ii) Berechne die größte Differenz (Distanz) zu einem Nachbarn in $N(x_0)$

$$\Delta(x_0) = \max_{N(x_0)} |x_0 - x_i|.$$

(iii) Ordne jedem Punkt in $N(x_0)$ ein Gewicht zu, wobei die Gewichtungsfunktion definiert ist durch

$$w(\tilde{u}) = w\left(\frac{|x_0 - x_i|}{\Delta(x_0)}\right),$$

wobei

$$w(\tilde{u}) = \begin{cases} (1-\tilde{u}^3)^3 & \text{für} \quad 0 \leq \tilde{u} < 1 \\ 0 & \text{sonst.} \end{cases}$$

(iv) Berechne die gewichtete KQ-Anpassung von y^* in $N(x_0)$, d.h.

$$\Sigma_{i=1}^k (y_i^* - f(x_i))^2 w(\tilde{u}_i) = Min,$$

und verwende als Glättungswerte für y^*

$$\hat{y}_0^* = \hat{f}(x_0).$$

Wenn für alle Beobachtungen $i = 1, \cdots, N$ und alle Regressoren $j = 1, \ldots, K$ die Schätzungen durchgeführt worden sind, folgt *Stufe 3*. Damit wird erneut mit $j = 1$ begonnen, jetzt auf einer höheren Stufe der Information über $f_2^{(2)}(x_2), \ldots, f_K^{(2)}(x_K)$. Auf jeder Stufe müssen die zuerst ermittelten $u_j^{(\cdot)}$ und folgend $f_j^{(\cdot)}$ mit weniger Information auskommen als die Späteren. Das Verfahren wird auf der m-ten Stufe abgebrochen, wenn

$$E(y - \beta_0 - \sum_{k=1}^{K} f_k^{(m)}(x_k))^2$$

nicht weiter reduziert werden kann. Dann steht für jeden Regressor x_k und jede Beobachtung $i = 1, \ldots, N$ ein geschätzter Wert zur Verfügung. Diese Werte beschreiben den Verlauf der Projektion von x_k auf y und lassen sich graphisch darstellen. Ein anderes Verfahren zur Anpassung der Daten an das additive Modell wird von Staniswalis und Severeni (2000) vorgeschlagen.

Ganz analog ist eine Schätzung eines verallgemeinerten additiven Modells mit diskreten endogenen Variablen möglich. Als Anfangsschätzung wird der OLS-Schätzer des linearen Wahrscheinlichkeitsmodells verwendet. Unter Ausnutzung folgender Definitionen und Beziehungen

$$y = \begin{cases} 1, & wenn \quad y^* > 0 \\ 0 & sonst \end{cases}$$

$$ln\frac{p_i}{1-p_i} = x_i'\beta + u_i$$

$$\eta_i^{(0)} \quad =: \quad x_i'\hat{\beta}, \qquad \mu_i^{(0)} =: \frac{exp(x_i'\hat{\beta})}{1+exp(x_i'\hat{\beta})}$$

kann die stufenweise Schätzung, zerlegt in Einzelschritte, erfolgen.

1. Stufe:

Schritt 1

$$y_{i1}^{(1)} =: x_i'\hat{\beta} + \frac{y_i - \mu_i^{(0)}}{\mu_i^{(0)}(1-\mu_i^{(0)})}$$

$y_{i1}^{(1)} = f(x_1)$ nichtparametrisch schätzen: $f^{(1)}(x_{i1})$.

Schritt 2

$$\begin{aligned} y_{i2}^{(1)} \quad &=: \quad (x_i'\hat{\beta} - \hat{\beta}_0 - x_{i1}\hat{\beta}_1 + f^{(1)}(x_{i1})) + \bar{y}_1^{(1)} + \frac{y_{i1}^{(1)} - \mu_{i1}^{(1)}}{\mu_{i1}^{(1)}(1-\mu_{i1}^{(1)})} \\ &=: \eta^{(1)}(x_{i1}) + \hat{\beta}_{01}^{(1)} + u_{i1}^{s(1)} \end{aligned}$$

wobei

$$\eta^{(1)}(x_{i1}) = x_i'\hat{\beta} - \hat{\beta}_0 - x_{i1}\hat{\beta}_1 + f^{(1)}(x_{i1})$$

$$\hat{\beta}_{01}^{(1)} = \bar{y}_1^{(1)} = \frac{1}{N}\sum_{i=1}^{N} y_{i1}^{(1)}$$

$$\mu_{i1}^{(1)} = \frac{exp(\eta^{(1)}(x_{i1}) + \hat{\beta}_{01}^{(1)})}{1+exp(\eta^{(1)}(x_{i1}) + \hat{\beta}_{01}^{(1)})}$$

$y_{i2}^{(1)} = f(x_2)$ nichtparametrisch schätzen: $f^{(1)}(x_{i2})$.

$\vdots$

Schritt K

$$y_{iK}^{(1)} =: (x_i'\hat{\beta} - \hat{\beta}_0 - \sum_{k=1}^{K-1} x_{ik}\hat{\beta}_k + \sum_{k=1}^{K-1} f^{(1)}(x_{ik})) + \hat{\beta}_{0;K-1}^{(1)} + \frac{y_{i;K-1}^{(1)} - \mu_{i;K-1}^{(1)}}{\mu_{i;K-1}^{(1)}(1-\mu_{i;K-1}^{(1)})}$$

$$=: \eta^{(1)}(x_{i;K-1}) + \hat{\beta}^{(1)}_{0;K-1} + u^{s(1)}_{i;K-1} + u^{s(1)}_{i,K-1}$$

$y^{(1)}_{iK} = f(x_K)$ nichtparametrisch schätzen: $f^{(1)}(x_{iK})$.

2. Stufe:

Schritt 1

$$y^{(2)}_{i1} =: \sum_{k=1}^{K} f^{(1)}(x_{ik}) + \hat{\beta}^{(1)}_{0K} + \frac{y^{(1)}_{iK} - \mu^{(1)}_{iK}}{\mu^{(1)}_{iK}(1-\mu^{(1)}_{iK})}$$

$y^{(2)}_{i1} = f(x_{i1})$ nichtparametrisch schätzen: $f^{(2)}(x_{i1})$.

Schritt 2

$$y^{(2)}_{i2} =: (\sum_{k=2}^{K} f^{(1)}(x_{ik}) + f^{(2)}(x_{i1})) + \hat{\beta}^{(2)}_{01} + \frac{y^{(2)}_{i1} - \mu^{(2)}_{i1}}{\mu^{(2)}_{i1}(1-\mu^{(2)}_{i1})} =: \eta^{(2)}(x_{i1}) + \hat{\beta}^{(2)}_{01} + u^{s(2)}_{i1}$$

$y^{(2)}_{i2} = f(x_{i2})$ nichtparametrisch schätzen: $f^{(2)}(x_{i2})$

$\vdots$

Schritt K

$$y^{(2)}_{iK} =: (f^{(1)}(x_{iK}) + \sum_{k=1}^{K-1} f^{(2)}(x_{ik})) + \hat{\beta}^{(2)}_{0;K-1} + \frac{y^{(2)}_{i;K-1} - \mu^{(2)}_{i;K-1}}{\mu^{(2)}_{i;K-1}(1-\mu^{(2)}_{i;K-1})} =: \eta^{(2)}(x_{i;K-1}) + \hat{\beta}^{(2)}_{0;K-1} + u^{s(2)}_{i;K-1}$$

$y^{(2)}_{iK} = f(x_{iK})$ nichtparametrisch schätzen: $f^{(2)}(x_{iK})$.

Entsprechend ist fortzufahren. Eine Verallgemeinerung dieses Ansatzes für ein multinomiales Modell entwickelt Abe (1999). Als Spezialfall lässt sich der bivariate Discrete-Choice-Ansatz im verallgemeinerten additiven gemischten Modell von Chen (2000) auffassen, das eine Mischung aus dem additiven Ansatz von Hastie und Tibshirani (1997) und den verallgemeinerten linearen Modellen des Typs McCullagh/Nelder (1983) darstellt. Horowitz (2001) entwickelt einen Ansatz im Rahmen verallgemeinerter additiver Modelle, wenn die Link-Funktion unbekannt ist

$$E(y|X = x, W = w) = g[\sum_{k=1}^{K} f_k(x^k), f_w(w)],$$

wobei $g[\cdot]$ und f_k unbekannte Funktionen sind $(k = 1, \cdots, K)$. W ist ein Zufallsvektor, der kontinuierliche und diskrete Komponenten enthalten kann. Während die f_k-Elemente additiv verknüpft sind, muss dies für f_w nicht gelten. Dieser Ansatz enthält das Single-Index-Modell und die verallgemeinerten additiven Modelle ebenso als Spezialfälle wie die verallgemeinerten multiplikativen Modelle. Im Rahmen eines triangulären Mehrgleichungsmodells kann der additive Ansatz eingesetzt werden, wie Newey, Powell und Vella (1999) aufzeigen.

Die Inferenz für additive Modelle ist bisher noch unterentwickelt. Einen Ansatzpunkt bietet die Erkenntnis, dass

(90) $$\hat{f}_k = W_k u^s_k$$

mit

$$V(\hat{f}_k) = W_k W_k' \sigma_k^2,$$

wobei

$$\hat{\sigma}_k^2 = \frac{\sum_{i=1}^N (u_{ik}^s - \hat{f}_k(x_{ik}))^2}{sp(I - W_k)}$$

ein linearer Glätter ist. Somit lässt sich ein Konfidenzband formulieren

$$(\hat{f}_{ik} - 2\hat{\sigma}_{\hat{f}_{ik}}; \hat{f}_{ik} + 2\hat{\sigma}_{\hat{f}_{ik}}).$$

(90) über alle Regressoren zusammengefasst, führt zu

$$\hat{f} \ = \begin{pmatrix} \hat{f}_1 \\ \vdots \\ \hat{f}_K \end{pmatrix} = \begin{pmatrix} W_1 & 0 & 0 & \cdots & 0 \\ 0 & W_2 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \\ 0 & 0 & 0 & \cdots & W_K \end{pmatrix} \cdot \begin{pmatrix} u_1^s \\ \vdots \\ u_K^s \end{pmatrix} =: W u^s.$$

Die gesamte Residuenquadratsumme entspricht

$$RSS = \sum_{i=1}^N (y_i - \hat{\beta}_0 - \sum_{k=1}^K \hat{f}_{ik})^2.$$

Als Test auf Signifikanz einzelner oder verschiedener Determinanten gleichzeitig kann

$$F = \frac{(RSS_1 - RSS_2)/(\gamma_2 - \gamma_1)}{RSS_2/(N - \gamma_2)} \sim F_{N-\gamma_2}^{\gamma_2 - \gamma_1}$$

verwendet werden, wobei $\gamma_\nu = sp(2W_\nu - W_\nu W_\nu')$; $\nu = 1, 2$. Die Indizes 1 und 2 stehen für restringiertes bzw. unrestringiertes Modell. Zur Gütebeurteilung des Gesamtmodells lässt sich die Deviance D heranziehen, definiert durch

$$D = -2[lnL(\hat{f}_1, ..., \hat{f}_K, y) - lnL(y, y)].$$

Eine Pseudo-Likelihood-Ratio-Teststatistik ist

$$\begin{aligned} PLRT &= D(\hat{\beta}, y) - D(\hat{f}_1, ..., \hat{f}_K, y) \\ &= D_R - D_U \sim \chi^2_{DF_U - DF_R}. \end{aligned}$$

Cai, Fan und Li (2000) stellen einen Goodness-of-Fit-Test vor, der auf einem nichtparametrischen Maximum-Likelihood-Ansatz basiert. Aufzudecken gilt es, ob gewisse Koeffizienten in einem additiven Modell konstant sind oder ob irgendwelche Kovariate im Modell statistisch signifikante Effekte ausüben. Die Verteilung der Teststatistik unter der Nullhypothese wird durch eine bedingte Bootstrapmethode geschätzt. Strukturelle Tests in additiven Modellen haben Härdle, Sperlich und Spokoiny (2001) entwickelt. Sie basieren auf Wavelet-Transformationen. Es geht darum zu überprüfen, ob einige additive Komponenten eine polynomiale Struktur besitzen, ohne die Struktur der restlichen Komponenten zu spezifizieren. Die zunächst nur für ein Modell mit einer nichtparametrischen Komponente entwickelten Tests lassen sich übertragen auf mehr als eine Komponente. Daneben kann mit diesem Ansatz auch auf Existenz von Interaktion verschiedener nichtparametrischer Terme geprüft werden, ob einzelne Regressoren überhaupt einen statistisch signifikanten Einfluss ausüben.

5.1.4 Empirische Studien

In der Zwischenzeit liegt eine Reihe ökonomischer Anwendungen mit nicht- oder semiparametrischen sowie additiven Modellen vor. So werden nichtparametrische Engel-Kurven von Blundell und Duncan (1998) geschätzt. Eine weitere Anwendung stammt von Horowitz und Härdle (1996), die die Wahrscheinlichkeit von Produktinnovation in deutschen Unternehmen der Investitionsgüterindustrie bestimmen. DiNardo und Tobias (2001, 22ff) demonstrieren die Arbeitsweise nichtparametrischer Schätzer in graphischer Darstellung für Einkommensfunktionen (log. Stundenlöhnen) in Abhängigkeit von intellektuellen Fähigkeiten A und Schulbildung S. Es können Effekte zwischen A und S verdeutlicht werden, die aus üblichen loglinearen Einkommensfunktionen nicht herauszuholen sind. Gerfin (1996) schätzt die Wahrscheinlichkeit eines Arbeitsangebotes mit Hilfe von Querschnittsdaten aus Deutschland und der Schweiz parametrisch, nicht- und semiparametrisch. Es zeigt sich, dass der parametrische Probitansatz für Deutschland abgelehnt wird, nicht jedoch für die Schweiz. Die geschätzten Koeffizienten variieren nicht systematisch zwischen den verschiedenen Methoden. Werwatz (1996) untersucht den Grad der Firmenspezifität der Lehrlingsausbildung in Deutschland anhand von Einkommensfunktionen für Betriebswechsler und für im Ausbildungsbetrieb verbliebene Beschäftigte. Um Selektionseffekte zu vermeiden, wird ein nichtparametrisches Verfahren zur Bestimmung der Selektionsgleichung angewandt. Semiparametrische Schätzungen für Selektionsmodelle von Einkommensfunktionen mit Hilfe der Daten des niederländischen sozioökonomischen Panels liefern Melenberg und van Soest (1993). Akkordlöhne und deren Effekte auf die Produktivität, ein anderer Aspekt der Arbeitsökonomik, werden von Shearer (1996) unter Verwendung semiparametrischer Ansätze analysiert. Die Migrationsneigung Ostdeutscher unter Verwendung der Daten des Sozio-ökonomischen Panels aus dem Jahre 1991 steht bei Müller (1997) und Burda, Härdle, Müller und Werwatz (1998) im Zentrum semiparametrischer Analyse. Chan und Randall (1997) nutzen einen semiparametrischen Ansatz für binäre Response-Modelle bei Untersuchungen über die Zahlungsbereitschaft zur Verbesserung der Umweltqualität. Anwendungen zu additiven Modellen sind für den Bereich betriebliche Weiterbildung und Mobilität (Hübler/König 1999, 1999a) sowie für den Zusammenhang zwischen Entlohnungsform und Marktmacht (Hübler 2001) zu finden. Thornton und Thomson (2001) nutzen additive Modelle bei der Analyse von Effekten, die vom Lernen aus Erfahrung mit gleichen und ähnlichen Produkten in Bezug auf die Produktivität ausgehen. Bianchi, Gudmundsson und Zoega (2001) betrachten das veränderte Arbeitsangebot in Abhängigkeit verschiedener Determinanten im Rahmen verallgemeinerter additiver Modelle. Ein inhaltlich breites Spektrum an Anwendungen mit nicht- und semiparametrischen Schätzungen findet sich in Fomby und Hill (2000).

5.2 Ansätze für Paneldaten

Soweit eine gepoolte Schätzung für Paneldaten mit nicht- und semiparametrischen Methoden durchgeführt werden soll, bedarf es keiner weiteren Ausführungen gegenüber den voranstehenden Abschnitten. Es ist lediglich vorher zu testen, ob das gepoolte Vorgehen angemessen ist. Parametrische Verfahren hierfür sind ausführlich bei Baltagi (2001) beschrieben. Einen nichtparametrischer Test für die Entscheidung, ob Poolbarkeit zulässig ist, präsentieren Baltagi, Hidalgo und Li (1996). Ob Individualeffekte existieren, lässt sich mit einem Test von Li und Hsiao (1996) prüfen. Individualeffekte lassen sich durch Fixed- oder Random-Effects-Ansätze modellieren.

5.2.1 Random-Effects-Modelle

Wird von folgendem einfachen nichtparametrischen Modell mit Individualeffekt α_i

(91) $$y_{it} = m_t(x_{it}) + (\alpha_i + \epsilon_{it}) = E(y_{it}|x_{it}) + u_{it} \qquad i = 1, \cdots, N \quad t = 1, \cdots, T$$

ausgegangen, so ist jedes Conditional-Moment-Verfahren zur Schätzung des nichtparametrischen Terms $m_t(x_{it})$ geeignet, wenn ein Random-Effects-Modell unterstellt wird. Die Verteilung des Individualeffektes wird als bekannt angenommen. Die α_i sollen identisch unabhängig verteilt und unabhängig von den Regressoren sein. Wegen der angenommenen unspezifischen Zeitabhängigkeit von $E(y_{it}|x_{it})$ ist getrennt für jede Welle zu schätzen. Wird Zeitinvarianz unterstellt, d.h. $m_t(x_{it}) = m(x_{it})$, dann kann das gepoolte Vorgehen für den nichtparametrischen Term gewählt werden. Ansonsten bietet sich ein lokal linearer Ansatz an

(92) $$y_{it} \approx m(x) + (x_{it} - x)'\beta(x) + \alpha_i + \epsilon_{it} =: \tilde{x}'_{it}\tilde{\beta} + \alpha_i + \epsilon_{it},$$

wobei $\tilde{x}' = (1, (x_{it} - x)')$ und $\tilde{\beta}' = (m(x), \beta(x)')$. Zur Bestimmung der nichtparametrischen Terme $m(x)$ und $\beta(x)$ kann das GLS-Vorgehen von Ullah und Roy (1998) gewählt werden, die der herkömmlichen Within-Transformation entspricht, die bei linearen Ansätzen üblicherweise verwendet wird. Dadurch wird α_i eliminiert.

Bei semiparametrisch partiell linearen Modellen für Paneldaten

(93) $$y_{it} = m(x_{it}) + z'_{it}\gamma + \alpha_i + \epsilon_{it}$$

kann bei Annahme von zeitinvarianten nichtparametrischen Termen dem Vorgehen bei Li und Stengos (1996) gefolgt werden. Grundgedanke ist, den nichtparametrischen Teil zu eliminieren, um anschließend den linearen Teil isoliert schätzen zu können, ein Vorgehen, das bereits bei Querschnittsdaten Verwendung findet und auf Robinson (1988) zurückgeht. Anders ausgedrückt, der Regressand ist die Differenz zwischen dem eigentlichen Regressanden und dessen Erwartungswert, der durch die nichtparametrischen Regressoren bedingt ist. Das ursprüngliche Modell wird um den Teil bereinigt, den die nichtlinearen Terme hervorrufen. Ein daraus resultierender OLS-Schätzer für den linearen Koeffizientenvektor ist konsistent. Der GLS-Schätzer ist, wie Li und Ullah (1998) zeigen, überlegen. Er erreicht die semiparametrische Effizienzgrenze. Vorher sind die bedingten Erwartungswerte zu bestimmen. Sie lassen sich durch ihre nichtparametrische Schätzung ersetzen. Eine geschickte Wahl der Gewichte mit Hilfe der geschätzten Dichtefunktion beseitigt den zufälligen Term der bedingten Erwartungswerte im Nenner des Nadaraya-Watson-Schätzers. Der ursprünglich für zeitinvariante nichtparametrische Terme eines gleichgewichtigen Panels entwickelte Schätzer lässt sich auf ein ungleichgewichtigtes Panel (König 1997) ebenso übertragen wie auf Modelle mit zeitvariablen nichtparametrischen Termen erweitern (König 2002). Wesentlich im zweiten Fall ist ein wellenspezifisches Vorgehen.

Eine Alternative zu dem Vorgehen von Li und Stengos entwickelt König (2002, 176ff). Die Vorteile liegen darin, dass die Besonderheiten der Panelstruktur besser berücksichtigt werden, also auch nichtparametrische Terme zeitvariabel sein können. Zudem werden einfachere Schätzer angeboten. Dem Paneldatenmodell müssen dann allerdings Restriktionen auferlegt werden. Übliche Differenzen- und Within-Schätzer, die aus rein linearen Modellen bekannt sind, lassen sich übertragen. Dies hat den Vorteil, dass sie von den sensitiv auf Bandweitenvariation reagierenden nichtparametrischen Schätzern unabhängig sind. Ein besseres Verhalten bei endlichen Stichproben ist somit zu erwarten. Nachteile sind bei asymptotischer Betrachtung möglich. Für

Anwendungen sind unendlich große Stichproben jedoch weniger bedeutsam. Das partiell lineare Modell kann bei zeitinvarianten nichtparametrischen Regressoren ($x_{it} = x_i$) als ein einfaches lineares Modell mit fixen Individualeffekten interpretiert werden

$$(94) \qquad y_{it} = z_{it}'\gamma + c_i + \epsilon_{it},$$

wobei $c_i = m(x_i) + \alpha_i$. Der Koeffizientenvektor γ kann so mit Hilfe der üblichen Differenzen- oder Within-Schätzer bestimmt werden, wenn die linearen Regressoren strikt exogen sind. Eine direkte Übertragung auf zeitvariable Regressoren führt zu verzerrten Schätzern. Der Bias verschwindet nur bei gleichem bedingten Erwartungswert des nichtparametrischen Terms über alle Perioden. Ansonsten lässt sich der Bias erheblich reduzieren, wenn dem interessanten Vorschlag von König (2002,182) gefolgt wird. Starke Verzerrungen ergeben sich immer dann, wenn die Werte zwischen zwei Perioden deutlich auseinander liegen. Daher werden solche Werte nur mit einem geringen Gewicht versehen, um die Verzerrung so weit wie möglich verschwinden zu lassen. Eine derartige Gewichtung erhält man durch Verwendung von Kernfunktionen. Der Schätzer für den Within-Fall lautet dann

$$(95)\ \hat{\gamma} = (\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{\tau=1}^{T} K_{(x_{it}-x_{i\tau})}(z_{it}-z_{i\tau})(z_{it}-z_{i\tau})')^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{\tau=1}^{T} K_{(x_{it}-x_{i\tau})}(z_{it}-z_{i\tau})(y_{it}-y_{i\tau}).$$

Dieser Schätzer ist konsistent und asymptotisch normalverteilt. Ein analoger Schätzer für den Differenzenfall ist problemlos zu bilden.

Zur Bestimmung der Verteilungsfunktion der Störgrößen eines Random-Effects-Modells ohne nichtparametrischen Term schlagen Horowitz und Markatou (1996) einen Kernschätzer vor. Die Schätzung ist getrennt für die empirischen Verteilungen von ϵ_{it} und $u_{it} = \alpha_i + \epsilon_{it}$, den gesamten Störterm, vorzunehmen ($\hat{u}_{it}, \hat{\epsilon}_{it}$). Die Isolierung von ϵ_{it} erfolgt, wie üblicherweise bei Paneldaten, durch Differenzenbildung

$$\hat{\epsilon}_{it} = (y_{it} - y_{i1}) - (z_{it} - z_{i1})'\hat{\gamma}.$$

Unter sehr schwachen Annahmen ist der Schätzer für die Dichtefunktionen von $\hat{u}_{it}$ und $\hat{\epsilon}_{it}$ konsistent.

5.2.2 Fixed-Effects-Modelle

Wird folgendes einfaches nichtparametrisches Paneldatenmodell mit fixen Individualeffekten

$$(96) \qquad y_{it} = m(x_{it}) + \alpha_i + \epsilon_{it}$$

formuliert, so taucht wegen der zugelassenen Abhängigkeit zwischen α_i und x_{it} gegenüber dem Random-Effects-Ansatz ein Problem auf. Der bedingte Erwartungswert von y_{it} entspricht nicht dem nichtparametrischen Term. Vielmehr gilt

$$(97) \qquad E(y_{it}|x_{it}) = m(x_{it}) + E(\alpha_i|x_{it}).$$

Der nichtparametrische Teil kann somit nicht mehr aus einem einfachen Conditional-Moment-Ansatz bestimmt werden. Die in linearen Modellen übliche Lösung über Differenzenbildung oder Within-Schätzer versagt hier. Der Individualeffekt wird damit zwar eliminiert. Der nichtparametrische Teil ist so aber nicht identifiziert. Über $y_{it} - y_{i,t-1}$ oder $y_{it} - \bar{y}_i$ bleiben $u_{it} - u_{i,t-1}$ bzw. $u_{it} - \bar{u}_i$ erhalten. Ullah und Roy (1998) gehen von der Taylorreihenentwicklung des nichtparametrischen Ausdruck aus

$$(98) \qquad y_{it} = m(x) + (x_{it} - x)' \frac{\partial m(x_{it})}{\partial x_{it}}|_{x_{it}=x} + \frac{1}{2}(x_{it} - x)' \frac{\partial^2 m(\tilde{x})}{\partial x \partial x'}(x_{it} - x) + \alpha_i + \epsilon_{it}$$

$$=: m(x) + (x_{it} - x)'\beta(x) + R_2(x_{it}, x) + \alpha_i + \epsilon_{it} =: m(x) + (x_{it} - x)'\beta(x) + \alpha_i + \tilde{\epsilon}_{x,it},$$

wobei $\tilde{x}$ zwischen x und x_{it} liegt. Es entsteht ein lokal lineares Modell. Über eine Within-Spezifikation

$$(99) \qquad y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)'\beta(x) + \epsilon_{x,it} - \bar{\epsilon}_{x,i}$$

soll eine Schätzung für $\beta(x)$ erreicht werden. Vorgeschlagen wird ein lokaler Within-Schätzer mit aus der Kernfunktion gebildeten Gewichten. Dieses Vorgehen führt aufgrund der Restterme $(E(\bar{R}_2(x_{i,x}|x_{it} = x) \neq 0)$ zu inkonsistenten Schätzungen. Gleiches gilt für einen analogen Differenzenschätzer (König 2002,60). Um bei einem lokal polynomialen Kleinste-Quadrate-Ansatz konsistent schätzen zu können, muss der Bias für $N \to \infty$ verschwinden. Dies gelingt König (2002,61ff) durch eine geeignete Gewichtung. Im Fall des Differenzenschätzers wird doppelt gewichtet. Ausgangspunkt bildet der Produktkern aus Periode t und t-1

$$(100) \qquad K(\frac{x_{it} - x}{h_N}, \frac{x_{i,t-1} - x}{h_N}) = K(\frac{x_{it} - x}{h_N}) \cdot K(\frac{x_{i,t-1} - x}{h_N}) =: K_{it} \cdot K_{i,t-1},$$

wobei h_N der Bandweite entspricht. Statt im Differenzenschätzer nur mit $K(\frac{x_{it}-x}{h_N})$ zu gewichten, wird mit dem Produkt der lokalen Kerne gewichtet

$$(101) \qquad \hat{\beta}(x)_{DIFF} = \{\sum_{i=1}^{N}\sum_{t=2}^{T} K_{it}K_{i,t-1}\Delta x_{it}\Delta x'_{it}\}^{-1} \sum_{i=1}^{N}\sum_{t=2}^{T} K_{it}K_{i,t-1}\Delta x_{it}\Delta y_{it},$$

wobei $\Delta x_{it} = x_{it} - x_{i,t-1}$ und $\Delta y_{it} = y_{it} - y_{i,t-1}$. Dieser Schätzer ist nicht nur konsistent, sondern auch asymptotisch normalverteilt mit einem Nullvektor als Erwartungswertvektor und einer asymptotischen Sandwich-Kovarianzmatrix. Auch für das Within-Modell ist eine geeignete Gewichtung derart möglich, dass der Schätzer konsistent ist. Hierfür leitet König (2002,68) den folgenden Schätzer ab

$$(102) \qquad \hat{\beta}(x)_{WITHIN} = \{\sum_{i=1}^{N}\sum_{t=1}^{T} K_{it}\tilde{x}_{it}\tilde{x}'_{it}\}^{-1} \sum_{i=1}^{N}\sum_{t=2}^{T} K_{it}\tilde{x}_{it}\tilde{y}_{it},$$

wobei

$$\tilde{x}_{it} = x_{it} - \frac{\sum_{t=1}^{T} K_{it}x_{it}}{\sum_{t=1}^{T} K_{it} + I[\sum_{t=1}^{T} K_{it} = 0]}.$$

Entsprechend ist $\tilde{y}_{it}$ definiert. $I[\cdot]$ beschreibt eine Indikatorfunktion, die den Wert Eins annimmt, falls die Bedingung in $[\cdot]$ erfüllt ist, andernfalls den Wert Null. Es lässt sich bei diesem Vorschlag eine gewisse Parallele zum Vorschlag von Blundell, Griffith und Windmeijer (2002) bei nicht strikt exogenen Variablen im Zähldatenmodell erkennen. Insofern liegt es nahe, $\tilde{x}_{it}$ als Quasi-Within-Variable zu bezeichnen. Die asymptotischen Varianzen des Within-Schätzers, die aus einer analogen Sandwich-Kovarianzmatrix wie beim Differenzenschätzer stammen, sind höchstens so groß wie die korrespondierenden Varianzen des Differenzenschätzers, falls lediglich zwei Perioden vorliegen. Bei $T \geq 3$ ist die Differenz der asymptotischen Kovarianzmatrizen $(V_{DIFF} - V_{WITHIN})$ positiv definit (König 2002,72).

Semiparametrisch partiell lineare Modelle mit fixen Individualeffekten lassen sich durch

$$(103) \qquad y_{it} = \tilde{m}(x) + x'_{it}\beta(x) + z'_{it}\gamma + \alpha_i + \tilde{\epsilon}_{it}$$

darstellen, wenn $\tilde{\epsilon}_{it} = \epsilon_{it} + R(x_{it}, x)$, $\tilde{m}(x) = m(x) - x'\beta(x)$. Der Individualterm α_i kann mit x_{it} und z_{it} korreliert sein und der nichtparametrische Term $m(x_{it})$ wird als Taylorreihe entwickelt. Das Problem einer Schätzung für den Parametervektor des linearen Teils (γ) besteht auch hier in dem bedingten Erwartungswert für y_{it}

$$(104) \qquad E(y_{it}|x_{it}) = \tilde{m}(x) + x'_{it}\beta(x) + E(z_{it}|x_{it})'\gamma + E(\alpha_i|x_{it}).$$

Differenzenbildung, d.h. $y_{it} - E(y_{it}|x_{it})$, eliminiert zwar den nichtparametrischen Term, nicht jedoch den Individualeffekt. Es muss also zunächst α_i in geeigneter Weise entfernt werden. Li und Stengos (1996) verwenden einen Differenzenschätzer für γ, bei dem mit den Nadaraya-Watson-Kernschätzern gewichtet wird

$$(105) \qquad \hat{\gamma}_{DIFF} = \{\sum_{i=1}^{N}\sum_{t=2}^{T} \Delta\hat{\tilde{z}}_{it}\Delta\hat{\tilde{z}}'_{it}\hat{\tilde{f}}(x_{it}, x_{i,t-1})^2\}^{-1} \sum_{i=1}^{N}\sum_{t=2}^{T} \Delta\hat{\tilde{z}}_{it}\Delta\hat{\tilde{y}}_{it}\hat{\tilde{f}}(x_{it}, x_{i,t-1})^2,$$

wobei $\hat{\tilde{f}}(x_{it}, x_{i,t-1})$ die Kerndichteschätzung und

$$\Delta\hat{\tilde{z}}_{it} = z_{it} - z_{i,t-1} - [\hat{E}(z_{it}|x_{it}, x_{i,t-1}) - \hat{E}(z_{i,t-1}|x_{it}, x_{i,t-1})].$$

Durch die erste Differenzenbildung verschwindet α_i, allerdings noch nicht der nichtparametrische Term. Die Differenz $\tilde{m}(x_{it}) - \tilde{m}(x_{i,t-1})$ entfällt erst, wenn die Differenz der Erwartungswerte abgezogen wird. Entsprechend ist $\Delta\hat{y}_{it}$ definiert. Falls $\Delta\hat{\tilde{z}}_{it} = z_{it} - z_{i,t-1} - [\hat{E}(z_{it}|x_{it}, x_{i,t-1}) - \hat{E}(z_{i,t-1}|x_{it}, x_{i,t-1})]$ und $\Delta\tilde{u}_{it}$ korreliert sind, müssen für $\Delta\tilde{z}_{it}$ entsprechende Instrumentalvariablen herangezogen werden. König (2002, S.215) schlägt einen alternativen Schätzer vor, der ohne die Kernschätzergewichte auskommt

$$(106) \qquad \hat{\gamma}_{DIFF1} = \{\sum_{i=1}^{N}\sum_{t=2}^{T} \Delta\hat{\tilde{z}}_{it}\Delta\hat{\tilde{z}}'_{it}\hat{I}_{it}\}^{-1} \sum_{i=1}^{N}\sum_{t=2}^{T} \Delta\hat{\tilde{z}}_{it}\Delta\hat{\tilde{y}}_{it}\hat{I}_{it},$$

wobei $\hat{I}_{it} = I[\hat{\tilde{m}}(x_{it}, x_{i,t-1}) > b_N]$ eine Trimmfunktion ist und für $N \to \infty$ gilt $b_N \to 0$. Die Idee zur Trimmfunktion findet sich bei Li/Lu/Ullah (1996)sowie Li/Ullah (1998). Auch zu (106) lässt sich ein Within-Schätzer formulieren

$$(107) \qquad \hat{\gamma}_{WITHIN} = \{\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{\tau=1}^{T}(z_{it} - z_{i\tau} - \hat{E}_{z_{it}-z_{i\tau}})(z_{it} - z_{i\tau} - \hat{E}_{z_{it}-z_{i\tau}})'\hat{I}_{it}\}^{-1}$$

$$\cdot \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{\tau=1}^{T}(z_{it} - z_{i\tau} - \hat{E}_{z_{it}-z_{i\tau}})(y_{it} - y_{i\tau} - \hat{E}_{y_{it}-y_{i\tau}})\hat{I}_{it}.$$

Ein weiterer Schätzer geht auf Berg, Li und Ullah (2000) zurück. Dieser ist allerdings inkonsistent, falls $x_{i\tau}$ und z_{it}, gegeben x_{it}, nicht unabhängig voneinander sind.

Für Paneldatenmodelle mit fixen Effekten und dichotom endogenen Variablen hat Manski (1975, 1985, 1987) nichtparametrische Maximum-Score-Schätzer eingeführt. Weitere Modelle und Schätzer liegen von Lee (1999a) und Honore/Lewbel (2002) vor. Einen Überblick über Tobitmodelle mit nichtparametrischen Komponenten, die den Standardfall zensierter endogener Variablen, Selektionsmodelle sowie zensierte Mehrgleichungsmodelle einschließen, geben Kyriazidou (1995,1997), Honore/Kyriazidou (2000) und entwickeln eigene Varianten. Zum Eliminieren des Individual- und des Sample-Selection-Effektes wird von der Differenzenmethode Gebrauch gemacht. Kyriazidou (1997) erzielt bei der Differenzenbildung zwischen Paaren von Beobachtungen für die Selektionseffekte dadurch Werte von nahezu Null , dass über die Gewichtung der Beobachtungen mit einer Kernfunktion Paare mit großer Differenz ein geringes Gewicht erhalten und umgekehrt.

5.2.3 Empirische Studien

Nicht- und semiparametrische Verfahren für Fixed- und Random-Effects-Modelle werden von König (2002) auf Arbeitsnachfragefunktionen unter Verwendung von Betriebsdaten aus dem Hannoveraner Firmenpanel angewandt. Es lassen sich aufgrund der nichtparametrischen Schätzungen kurzfristig komplementäre und längerfristig eher substitutive Beziehungen zwischen Material und Kapital ausmachen. Die partiell linearen Schätzungen bei fixen Individualeffekten zeigen eine deutlich stärkere Nähe zu den üblichen Within- als zu den OLS-Schätzern. Die Berücksichtigung von Interaktionen zwischen Zeit- und Individualeffekten führt interessanterweise kaum zu Veränderungen gegenüber der reinen Individualeffektbetrachtung. Dies Ergebnis spricht dafür, dass in dem relativ kurzen Zeitraum von vier Jahren keine ausgesprochenen Zeiteffekte wirksam werden. In Hübler/König (1998) werden semiparametrisch partiell lineare Lohngleichungen unter anderem in Abhängigkeit von Marktmacht und Betriebsgröße geschätzt. Poolbarkeit der Daten wird nicht abgelehnt, wohl aber die These, dass keine Individualeffekte vorliegen. Ein geringer Marktanteil ist mit niedrigen Löhnen verbunden und aus hohen Marktanteilen folgen hohe Löhne. Die positiven Lohneffekte sind in Betrieben mittlerer Größe am stärksten ausgeprägt, ein Ergebnis, das erst durch die semiparametrische Schätzung aufgedeckt werden konnte. Horowitz und Markatou (1996) zeigen den Nutzen ihres nichtparametrischen Ansatzes zur Bestimmung der Verteilung des Individualeffektes α_i und der Reststörgröße ϵ_{it} anhand von Einkommensfunktionen für Individuen auf. Es zeigt sich, dass bei Annahme normalverteilter Reststörgrößen die Wahrscheinlichkeit, dass eine Person mit niedrigem Einkommen in Zukunft ein hohes Einkommen erzielt, erheblich überschätzt wird.

6 Schlussbemerkung

Dieser Beitrag stellt einige wesentliche Entwicklungen in den Methoden der Mikroökonometrie für die letzten 10 Jahre dar. Ausgehend von den Kernbereichen der Mikroökonometrie werden insbesondere drei Themen behandelt, die neue Impulse erfahren haben: Instrumentalvariablenschätzer bei schwachen Instrumenten, parametrische Paneldatenanalyse bei Linked-Employer-Employee-Daten (LEEP) sowie nicht- und semiparametrische Schätzer bei Querschnitts- und Paneldaten. Versucht man Verbindungen zwischen diesen Gebieten herzustellen, so ist zu erwarten, dass in Zukunft LEEP-Modelle auch unter dem Blickwinkel schwacher Instrumente betrachtet werden. Außerdem dürften semi- und nichtparametrische Schätzverfahren Eingang in die Analyse dieses Modelltyps finden. Für die Evaluation von Maßnahmeneffekte bieten

LEEP-Modelle ein breites Anwendungsfeld, wenn es darum geht, die direkten und indirekten Wirkungen betrieblicher und staatlicher Maßnahmen auf Beschäftigte und Unternehmen zu erfassen. Verstärkt sollten Instrumentalvariablen auch in nichtlinearen Modellen zum Einsatz kommen. Wünschenswert wäre parametrische und nichtparametrische Ansätze nicht nur als Alternativen zu betrachten, sondern auch als Ergänzung. Erstere können Hinweise liefern, wie letztere zu parametrisieren sind. Nichtparametrische Schätzer könnten verstärkt für die Analyse kurzfristiger Bewegungen von Bedeutung sein, während parametrische Schätzer mehr für langfristige gleichgewichtige ökonomische Betrachtungen herangezogen werden könnten. Zu wenig Aufmerksamkeit wird bisher der optimalen Aufteilung von Zeit- und Geldbudgets, der Einbeziehung von Zielfunktionen unter Nebenbedingungen sowie der Qualität der Daten und der Auswahl der Stichproben geschenkt. Hier bieten sich wesentliche Forschungsfelder im Rahmen der Mikroökonometrie an.

Literaturverzeichnis

Abe, M. (1999): A Generalized Additive Model for Discrete-Choice Data, Journal of Business & Economic Statistics 17, 271-284.

Abowd, J.M., Creecy, R.H. und F. Kramarz (2002): Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data, Cornell University Working Paper.

Abowd, J.M. und F. Kramarz (1999): The Analysis of Labor Markets Using Matched Employer-Employee Data, in: O. Ashefelter und D. Card (eds.), Handbook of Labor Economics, Vol. 3B, Elsevier: Amsterdam, 2629-2710.

Abowd, J.M. und F. Kramarz (1999a): Econometric Analyses of Linked Employer-Employee Data, Labour Economics 6, 53-74.

Abowd, J.M., Kramarz, F. und D.N. Margolis (1999): High Wage Workers and High Wage Firms, Econometrica 67, 251-333.

Ahn, H. (1997): Semiparametric Estimation of a Single-Index Model with Nonparametrically Generated Regressors, Econometric Theory 13, 3-31.

Ait-Sahalia, Y., Bickel, P.J. und T.M. Stoker (2001): Goodness-of-fit Tests for Kernel Regression with an Application to Option Implied Volatilies, Journal of Econometrics 105, 363 - 412.

Altonji,J. und L. Segal (1996): Small Sample Bias in GMM Estimation of Covariance Structure, Journal of Business and Economics Statistics 14, 353-366.

Andrews, D.W.K. (1999): Consistent Moment Selection Procedures for Generalized Method of Moments Estimation, Econometrica 67, 543-564.

Angrist, J.D., Imbens, G.W. und A.B. Krueger (1999): Jackknife Instrumental Variables Estimation, Journal of Applied Econometrics 14, 57-87.

Angrist,J.D. und A.B.Krueger (1991): Does Compulsory School Attendance Affect Schooling and Earnings, Quarterly Journal of Economics 106, 979-1014.

Angrist, J.D. und A.B. Krueger (1995): Split Sample Instrumental Variables Estimates of the Return to Schooling, Journal of Business and Economics Statistics 13, 225-235.

Angrist, J.D. und A.B. Krueger (1999): Empirical Strategies in Labor Economics, in: O. Ashenfelter und D. Card (eds.), Handbook of Labor Economics, Vol. 3B, Elsevier: Amsterdam, 1277 - 1366.

Angrist, J.D. und A.B. Krueger (2001): Instrumental Variables and the Search for Identification: From Supply andd Demand to Natural Experiments, Journal of Economic Perspectives 15, 69-86.

Angrist, J.D. und V. Lavy (1999): Using Maidonides Rule to Estimate the Effects of Class Size on Scholastic Achievement, Quarterly Journal of Economics 114, 533 - 575.

Arellano, M. (2002): Panel Econometrics, Oxford (forthcoming).

Arellano, M. und B. Honore (2001): Panel Data Models: Some Recent Developments, in: J.J. Heckman und E. Leamer (eds.), Handbook of Econometrics, Vol. 5, Elsevier: Amsterdam, 3229-3296.

Baltagi, B.H. (2001): Econometric Analysis of Panel Data, 2nd ed., Wiley: New York.

Baltagi, B.H., Hidalgo, J. und Q. Li (1996): A Nonparametric Test for Poolability Using Panel Data, Journal of Econometrics 75, 345-367.

Bellmann, L., Bender, S. und A. Kölling (2000): Aufbau eines Employer-Employee-Datensatzes aus IAB-Betriebspanel und Beschäftigtenstatistik der Bundesanstalt für Arbeit, Nürnberg, mimeo.

Bellmann, L.und U. Blien (2001): Wage Curve Analyses with Establishment Data from Western Germany, Industrial and Labor Relations Review 54, 851-863.

Berg, M.D., Li, Q. und A. Ullah (2000): Instrumental Variable Estimation of Semiparametric Dynamic Panel Data Models: Monte Carlo Results on Several New and Existing Estimators, in: B.H. Baltagi (ed.), Advances in Econometrics: Non-Stationary Panels, Panel Cointegration, and Dynamic Panels, Vol. 15, J.A.I.: New York, 297 - 315.

Bianchi, M., Gudmundsson, B.R. und G. Zoega (2001): Iceland's Natural Experiments in Supply-Side Economics, American Economic Review 91, 1564-1579.

Blundell, R. (2002): Dynamic Panel Data Methods and Practice, Allgemeines Statistisches Archiv 86, 145 - 162.

Blundell, R. und S. Bond (1998): Initial Conditions and Moment Restrictions in Dynamic Panel Data Models, Journal of Econometrics 87, 115 - 143.

Blundell, R., Griffith, R. und F. Windmeijer (2002): Individual Effects and Dynamics in Count Data Models, Journal of Econometrics 108, 113-131.

Blundell, R. und A. Duncan (1998): Kernel Regression in Empirical Microeconomics, Journal of Human Resources 33, 62-87.

Bond, S., Bowsher, C. und F. Windmeijer (2001): Criterion-based Inference for GMM in Autoregressive Panel Data Models, Economics Letters 73, 379-388.

Bowden, R.J. und D.A. Turkington (1984): Instrumental Variables, Cambridge University Press: Cambridge.

Börsch-Supan, A. (1994): Simulationsmethoden für die Analyse qualitativer Daten, Allgemeines Statistisches Archiv 78, 20-39.

Börsch-Supan, A. und V.A. Hajivassiliou (1993): Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models, Journal of Econometrics 58, 347-368

Bound, J., Jaeger, D.A. und R.M. Baker (1995): Problems With Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable Is Weak, Journal of the American Statistical Association 90, 443-450.

Breitung, J. (1992): Dynamische Modelle für die Paneldatenanalyse, Haag+Herchen: Frankfurt.

Breitung, J. (2000): The Local Power of Some Unit Root Tests for Panel Data, in: B.H. Baltagi (ed.), Nonstationary Panels, Panel Cointragtion, and Dynamic Panel, JAI: Amserdam, 161-177.

Buja, A., Hastie, T.J. und R.J. Tibshirani (1989): Linear Smoothers and Additive Models, Annals of Statistics 17, 453-555.

Burda, M.C., Härdle, W., Müller, M. und A. Werwatz (1998): Semiparametric Analysis of German East-West Migration Intentions: Facts and Theory, Journal of Applied Econometrics 13, 525-542.

Cameron, A.C. und P.K. Trivedi (1998): Regression Analysis of Count Data, Cambridge University Press: Cambridge.

Card, D. (2001): Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems, Econometrica 69, 1127-1160.

Carrasco, R. (2001): Binary Choice With Binary Endogenous Regressors in Panel Data: Estimating the Effect of Fertility on Female Labor Participation, Journal of Business & and Economic Statistics 19, 385-394.

Chay, K.Y. und B.E. Honore (1998): Estimation of Semiparametric Censored Regression Models, Journal of Human Resources 33, 4-38.

Chay, K.Y. und J.L. Powell: Semiparametric Censored Regression Models ,Journal of Economic Perspectives 15, 29-42.

Chen, C. (2000): Generalized Additive Mixed Modells, Communications in Statistics - Theory and Methods 29, 1257-1271.

Chen, H. und A. Randall (1997): Semi-nonparametric Estimation of Binary Response Models With an Application to Natural Resource Valuation, Journal of Econometrics 76, 323 - 340.

Chen, S. (1997): Semiparametric Estimation of the Type-3 Tobit Model, Journal of Econometrics 80, 1-34.

Chib, S. (2001): Markov Chain Monte Carlo Methods: Computation and Inference, in Econometrics, in: J.J. Heckman und E. Leamer (eds.), Handbook of Econometrics, Vol. 5, Elsevier: Amsterdam, 3569-3649.

Chib, S. und E. Greenberg (1995): Understanding the Metropolis-Hastings Algorithm, American Statistician 49, 327-335.

Chip, S., Greenberg, E. und R. Winkelmann (1998): Posterior Simulation and Bayes Factors in Panel Count Data Models, Journal of Econometrics 86, 33 - 54.

Czado, C. (2000): Multivariate Regression Analysis of Panel Data with Binary Outcomes Applied to Unemployment Data, Statistical Papers 41, 281 - 304.

Diggle, P.J., Liang, K.-Y. und S.L. Zeger (1994): Analysis of Longitudinal Data, Clarendon Press: Oxford.

DiNardo, J. und J.L. Tobias (2001): Nonparametric Density and Regression Estimation, Journal of Economic Perspectives 15, 11-28.

Donald, S.G. und W.K. Newey (2001): Choosing the Number of Instruments, Econometrica 69, 1161-1191.

Dustmann, C. und M.E. Rochina-Barrachina (2000): Selection Correction in Panel Data Models: An Application to Labour Supply and Wages, IZA-DP 162.

Fan, Y. und Q. Li (1996): Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms, Econometrica 64, 865-891.

Fiebig, D.G., Bartels, R. und W. Krämer (1996): The Frisch-Waugh Theorem and Generalized Least Squares, Econometric Reviews 15, 431-443.

Fomby, T.B. und R.C. Hill (2000): Applying Kernel and Nonparametric Estimation to Economic Topics, Advances in Econometrics, JAI Press: Stamford.

Gerfin, M. (1996): Parametric and Semiparametric Estimation of the Binary Response Model of Labour Market Participation, Empirical Economics 11, 321-339.

Geweke, J. und M. Keane (2001): Computationally Intensive Methods for Integration in Econometrics, in: J.J. Heckman und E. Leamer (eds.), Handbook of Econometrics, Vol. 5, Elsevier: Amsterdam, 3465-3568.

Gilks, W.R., Richardson, S. und D.J. Spiegelhalter (1997) Markov Chain Monte Carlo in Practice, Chapman & Hall:London.

Goldstein, H.(1995): Multilevel Statistical Models, 2nd. ed., Edward Arnold: London.

Gourieroux, C. und A. Monfort (1993): Pseudo-Likelihood Methods, in: Maddala, G.S., C.R. Rao, H.D. Vinod (eds.), Handbook of Statistics 11, North-Holland: Amsterdam, 335-362.

Gourieroux, C. und A. Monfort (1996): Simulation-Based Econometric Methods, Oxford University Press: Oxford.

Gourieroux, C., Monfort, A. und C. Tenreiro (2000): Kernel-M-estimators and Functional Residual Plots, in: J. Krishnakumar und E. Ronchetti (eds.), Panel Data Econometrics: Future Directions, Elsevier: Amsterdam, 235-275.

Goux, D. und E. Maurin (1999): Persistence of Interindustry Wage Differentials: A Reexamination Using Matched Worker-Firm Panel Data, Journal of Labor Economics 17, 492-533.

Groshen, E. (1996): American Employer Salary Surveys and Labor Economics Research: Issues and Contributions, Annales d'Economie et de Statistique 41/42, 413-442.

Gurmu, S., Rilstone, P. und S. Stern (1999): Semiparametric Estimation of Count Regression Models, Journal of Econometrics 88, 123-150.

Hahn, J. und J. Hausman (2002): A New Specification Test for the Validity of Instrumental Variables, Econometrica 70, 163-189.

Hamermesh, D.S. (1999): LEEping into the Future of Labor Economics: the Research Potential of Linking Employer and Employee Data, Labour Economics 6, 25-41.

Han, C. und P. Schmidt (2001): The Asymptotic Distribution of the Instrumental Variable Estimator When the Instruments Are Not Correlated With the Regressors, Economics Leters 74, 61 -66.

Hansen, B.E. (1999): Threshold Effects in Non-Dynamic Panels: Estimation, Testing, and Inference, Journal of Econometrics 93, 345-368.

Hansen, L.P. (1982): Large Sample Properties of Generalized Methods of Moments Estimators, Econometrica 50, 1029 - 1054.

Härdle, W. (1990): Applied Nonparametric Regression, Cambridge University Press: Cambridge.

Härdle, W., Liang, H. und J. Gao (2000): Partially Linear Models, Physica: Heidelberg.

Härdle, W., Sperlich, S. und V. Spokoiny (2001): Structural Tests in Additive Regression, Journal of the American Statistical Association 96, 1333-1347.

Hastie, T.J.und R.J. Tibshirani (1986): Generalized Additive Models, Statistical Science 1, 297-318.

Hastie, T.J. und R.J. Tibshirani (1997): Generalized Additive Models, Chapman and Hall: London.

Hausman, J.A. (2001): Microeconometrics, Journal of Econometrics 100, 33-35.

Hausman, J.A., Hall, B.H. und Z. Griliches (1984): Econometric Models for Count Data with an Application to the Patents-R and D Relationship, Econometrica 52,909-938.

Hausman, J.A. und D.A. Wise (1979): Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment, Econometrica 47, 455-473.

Heckman, J.J. (1974): Shadow Prices, Market Wages, and labor Supply, Econometrica 42, 679-694.

Heckman, J.J. und E. Leamer, eds. (2001): Handbook of Econometrics, Vol. 5, Elsevier: Amsterdam.

Hildreth, C. und J. Lu (1960): Demand Relations With Autocorrelated Disturbances, Technical Bulletin no. 276, Michigan State University, Agriculture Experiment Station.

Hildreth, A.K. und S. Pudney (1999): Econometric Issues in the Analysis of Linked Cross-section Employer-Worker Surveys, in: J. Haltiwanger et al. (eds.), The Creation and Analysis of Employer-Employee Matched Data, North-Holland: Amsterdam, 461-488.

Honore, B.E. und E. Kyriazidou (2000): Estimation of Tobit-Type Models with Individual Specific Effects, Econometric Review 19, 341-366.

Honore, B.E., Kyriazidou, E. und C. Udry (1997): Estimation od Type-3 Tobit Models Using Symmetric Trimming and Pairwise Comparisons, Journal of Econometrics 76, 107-128.

Honore, B.E. und A. Lewbel (2002): Semiparametric Binary Choice Panel Data Models without Strictly Exogeneous Regressors, forthcoming in Econometrica.

Horowitz, J.L. (1992): A Smoothed Maximum Likelihood Estimator for the Binary Response Model, Econometrica 60, 505-531.

Horowitz, J.L. (1998): Semiparametric Methods in Econometrics, Springer: New York.

Horowitz, J.L. (2001): Nonparametric Estimation of a Generalized Additive Model With an Unknown Link Function, Econometrica 69, 499-513.

Horowitz, J.L. und W. Härdle (1996): Direct Semiparametric Estimation of Single-Index Models with Discrete Covariates, Journal of the American Statistical Association 91, 1632-1640.

Horowitz, J.L. und M. Markatou (1996): Semiparametric Estimation of Regression Models for Panel Data, Review of Economic Studies 63, 145-168.

Horowitz, J.L. und N.E. Savin (2001): Binary Response Models: Logits, Probits and Semiparametrics, Journal of Economic Perspectives 15, 43-56.

Hsiao, C., Pesaran, M.H. und A.K. Tahmiscioglu (2002): Maximum Likelihood Estimation of Fixed Effects Dynamic panel Data Models Covering Short Time Periods, Journal of Econometrics 109, 107-150.

Hsiao, C. (1986): Analysis of Panel Data, Cambridge University Press: Cambridge.

Hübler, O. (1989): Ökonometrie, Gustav Fischer: Stuttgart.

Hübler, O. (1990): Lineare Paneldatenmodelle mit alternativer Störgrößenstruktur, in: G. Nakhaeizadeh und K.-H. Vollmer, Neuere Entwicklungen in der Angewandten Ökonometrie, Physica-Verlag: Heidelberg, 65-99.

Hübler, O. (1990a): Messung von Diskriminierung durch direkte und inverse Regression, Allgemeines Statistisches Archiv 74, 315-335.

Hübler, O. (2001): Payment Schemes and Market Shares - An Econometric Study with Establishment Data, in: R. Friedmann, L. Knüppel und H. Lütkepohl (eds.), Econometric Studies, A Festschrift in Honour of Joachim Frohn, Lit Verlag: Münster, 295-315.

Hübler, O. (2001a): Evaluation of Policy Interventions: Measurement and Problems, Allgemeines Statistisches Archiv 85, 103-126.

Hübler, O. und A. König (1998): Produktmarkteinflüsse, Renten und Löhne,in: K. Gerlach, O. Hübler und W. Meyer (Hrsg.), Ökonomische Analysen betrieblicher Strukturen und Entwicklungen. Das Hannoveraner Firmenpanel, Campus: Frankfurt, 119-143.

Hübler, O. und A. König (1999): Betriebliche Weiterbildung, Mobilität und Beschäftigungsdynamik, Jahrbücher für Nationalökonomie und Statistik 219, 165-193.

Hübler, O. und A. König (1999a): Verstärkt Weiterbildung die Betriebsbindung oder die Flexibilität der Beschäftigten?In: L. Bellmann und V. Steiner (Hrsg.), Panelanalysen zu Lohnstruktur, Qualifikation und Beschäftigungsdynamik, BeitrAB 229, 263-312.

Im, K.S., Ahn, S.C., Schmidt, P. und J.M. Wooldridge (1999): Efficient Estimation of Panel Data Models With Strictly Exgogenous Explanatory Variables, Journal of Econometrics 93, 177-201.

Imbens, G.W., Spady, R.H. und P. Johnson (1998): Information Theoretic Approaches to Inference in Moment Conditioned Models, Econometrica 66, 333-357.

Inkmann, J. (2001): Conditional Moment Estimation of Nonlinear Equation Models, Springer: Berlin.

Janz, N. (1997): Ökonometrische Paneldatenanalysen des Investitionsverhaltens deutscher Aktiengesellschaften, Nomos Verlagsgesellschaft: Baden-Baden.

Judge, G.G., Griffiths, W.E., Hill, R.C., Lütkepohl, H. und T.-C. Lee (1985): The Theory and Practice of Econometrics, 2nd. ed., Wiley & Sons: New York.

Kaltenborn, U. (1997): Die Anwendung simulativer Schätzverfahren für Discrete-Choice-Modelle am Beispiel von Daten des Sozioökonomischen Panels, Disseratation FU Berlin.

Ke, C: und Y. Wang (2001): Semiparametric Nonlinear Mixed-Effects Models and Their Applications, Journal of the American Statistical Association 96, 1272-1281.

Klein, R.L. und R.H. Spady (1993): An Efficient Semiparametric Estimator for Binary Response Models, Econometrica 61, 387-421.

Kniesner, T.J. und Q. Li (1996): Semiparametric Panel Data Models with Dynamic Adjustment: Theoretical Considerations and an Application to Labor Supply, mimeo.

König, A. (1997): Schätzen und Testen in semiparametrischen partiell linearen Modellen für die Paneldatenanalyse, Universität Hannover, Diskussionspapier Nr. 208.

König, A. (2002): Nichtparametrische und semiparametrische Schätzverfahren für die Paneldatenanalyse, Lit-Verlag: Münster.

König, H. und M. Lechner (1994): Some Recent Developments in Microeconometrics, Swiss Journal of Economics and Statistics 130, 299 - 331.

Kyriazidou, E. (1995): Essays in Estimation and Testing of Econometric Models, Dissertation, Evanston (Illinois).

Kyriazidou, E. (1997): Estimation of a Panel Data Sample Selection Model, Econometrica 65, 1335-1364.

Lavergne, P. und Q.H. Vuong (1996): Nonparametric Selection of Regressors: The Nonnested Case, Econometrica 64, 207-219.

Lechner, M. (2002): Eine Übersicht über gängige Modelle der Panelökonometrie und ihre kausale Interpretation, erscheint in: Allgemeines Statistisches Archiv.

Lee, L.F. (1982): Some Approaches to the Correction of Selectivity Bias, Review of Economic Studies 49, 355-372.

Lee, M.-J. (1996): Methods of Moments and Semiparametric Econometrics for Limited Dependent Variable Models, Springer: New York.

Lee, M.-J.(1999): Nonparametric Estimation and Test for Quadrant Correlation in Multivariate Binary Response Models, Econometric Reviews 18, 387-415.

Lee, M.-J. (1999a): A Root-N Consistent Semiparametric Estimator for Related-Effect Binary Response Panel Data, Econometrica 67, 427-433.

Lewbel, A. (1997): Semiparametric Estimation of Location and Other Discrete Choice Moments, Econometric Theory 13, 32-51.

Li, D. und T. Stengos (2002): The Partially Linear Regression Model: Monte Carlo Evidence from the Projection Pursuit Regression Approach, Economics Letters 75, 11 - 16.

Li, Q und C. Hsiao (1998): Testing Serial Correlation in Semiparametric Panel Data Models, mimeo, Department of Economics, Journal of Econometrics 87, 207 - 237.

Li, Q. und Stengos (1996): A Semi-Parametric Non-Nested Test in a Dynamic Panel Data Model, Economics Letters 49, 1-6.

Li, Q. und A. Ullah (1998): Estimating Partially Linear Panel Data Models With One-Way Error Components, Econometric Review 17, 145-166.

Li, Q. und S. Wang (1998): A Simple Consistent Bootstrap Test for a Parametric Regression Function, mimeo, Department of Economics, Journal of Econometrics 87, 145 - 165.

Longford, N.T. (1993): Random Coefficient Models, Clarendon Press: Oxford.

Manski, C.F.(1975): Maximum Score Estimation of the Stochastic Utility Model of Choice, Journal of Econometrics 3, 205-228.

Manski, C.F.(1985): Semiparametric Analysis of Discrete response: Asymptotic Properties of the maximum Score Estimator, Journal of Econometrics 27, 313-333.

Manski, C.F.(1987): Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data, Econometrica 55, 357-362.

Matzkin, R.L. (1992): Nonparametric and Distribution-Free Estimation of the Binary Threshold Crossing and the Binary Models, Econometrica 60., 239-270.

McCullagh, P. und J.A. Nelder (1983): Generalized Linear Models, Chapman & Hall: London.

Melenberg, B. und A. van Soest (1993): Semiparametric Estimation of the Sample Selection Model, CentER for Economic research, DP No. 9334.

Moffitt, R.A. (1999): New developments in Econometric Methods for Labor Market Analysis, in: O. Ashenfelter und D. Card (eds.), Handbook of Labor Economics, Vol. 3B, Elsevier: Amsterdam, 1367 - 1397.

Mühleisen, M. (1994): Human Capital Decay and Persistence. A Simulation Approach to German Unemployment, Campus: Frankfurt/M.

Müller, M. (1997): Computer-assisted Generalized Partial Linear Models, Humboldt-Universität, SFB 373, Discussion Paper No. 48/1997.

Nelson, C.R. und R. Startz (1990) : Some further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator, Econometrica 58, 967-976.

Nevo, A. (2002): Sample Selection and Information-theoretic Alternatives to GMM, Journal of Economtrics 107, 149-157.

Newey, W.K., Powell, J.L. und F. Vella (1999): Nonparametric Estimation of Triangular Simultanous Equations Models, Econometrica 67, 565-603.

Nijman, T. und M. Verbeek (1992): Nonresponse in Panel Data: The Impact on Estimates of a Life Cycle Consumption Function, Journal of Applied Econometrics 7, 243-257.

Pagan, A. und A. Ullah (1999): Nonparametric Analysis, Cambridge University Press: Cambridge.

Pesaran, M.H. und R.J. Smith (1994): A Genralized R^2 Criterion for Regression Models Estimated by the Instrumental Variables Method, Econometrica 62, 705-710.

Powell, J.L., Stocker, J.H. und T.M. Stoker (1989): Semiparametric Estimation of Index Coefficients, Econometrica 57, 1403 1430.

Robert, C.P. und G. Casella (1997): Monte Carlo Statistical Methods, Springer: New York - Berlin.

Robinson, P.M. (1988): Root-N-Consistent Semiparametric Regression, Econometrica 56, 931-954.

Rochina-Barrachina, M.E. (1997): A New Estimator for Panel Data Sample Selection Models, Thesis Dissertation UCL.

Ronning, G. (1991): Mikroökonometrie, Springer: Berlin.

Ronning, G. (2002): Estimation of Discrete Choice Models with Minimal Variation of Alternative-specific Variables, Economtric Review 21, 135-146.

Rosenzweig, M.R. und K.I. Wolpin (2000): Natural'Natural Experiments' in Economics, Journal of Economic Literature 38, 827-874.

Roy, A.D. (1951): Some Thoughts on the Distribution of Earnings, Oxford Economic Papers 3, 135-146.

Searle, S.R. (1971): Linear Models, Wiley: New York.

Shao, J. (1999): Mathematical Statistics, Springer: New York, Berlin.

Shearer, B. (1996): Piece-Rates, Principal Agent Models, and Productivity Profiles, Parametric and Semi-Parametric Evidence from Payroll Records, Journal of Human Resources 31, 275 - 303.

Staiger, D. und J.H. Stock (1997): Instruemntal Variables Regression with Weak Instruments, Econometrica 65, 557 - 586.

Staniswalis, J.G. und T.A. Severeni (2000): Fitting the Additive Model by Recursion on Dimension, Communications in Statistics - Simulations - 29(3), 689-701.

Stephan, G. (2001): Firmenlohndifferenziale, Campus: Frankfurt/New York.

Stern, S. (1997): Simulation-Based Estimation, Journal of Economic Literature 35, 2006-2039.

Stock,J.H. und J.H. Wright (2000): GMM with Weak Idenitification, Econometrica 68, 1055-1096.

Thornton, R.A. und P. Thomson (2001): Learning from Experience and Learning from Others: An Exploration of Learning and Spillovers in Wartime Shipbuilding, American Economic Reviev 91, 1350-1368.

Ullah, A. und N. Roy (1998): Nonparametric and Semiparametric Econometrics of Panel Data, in: A. Ullah und D.E.A. Giles (eds.), Handbook of Applied Economics, Marcel Dekker: New York, 579-604.

Ullah, A. und H.D. Vinod (1993): General Nonparametric Regression Estimation and Testing in Econometrics, in: G.S. Maddala, C.R. Rao, H.D. Vinod (eds.), Handbook of Statistics 11, North-Holland: Amsterdam, 85-116.

Vella, F. (1998): Estimating Models with Sample Selection Bias: A Survey, Journal of Human Resources 33, 127-169.

Vella, F. und M. Verbeek (1999): Estimating and Interpreting Models with Endogenous Treatment Effects, Journal of Business and Economic Statistics 17, 473 - 478.

Verbeek, M. (1991): The Design of Panel Surveys and the Treatment of Missing Observations, Dissertation University of Brabant.

Verbeek, M. und T. Nijman (1992): Testing for Selectivity Bias in Panel Data Models, International Economic Review 33, 681-703.

Verbeke, G. und G. Molenberghs (2000): Linear Mixed Models for Longitudinal Data, Springer: New York.

Vijverberg, W.P.M. (2000): A Family That Nests Probit and Logit, IZA-DP No. 222.

Wang, J. und E. Zivot (1998): Inference on Structural Parameters in Instrumental Variables Regression With Weak Instruments, Econometrica 66, 1389-1404.

Weeks, M. und C. Orme (1998): The Statistical Relationship Between Bivariate and Multinomial Choice Models, mimeo.

Werwatz, A. (1996): Wie firmenspezifisch ist die deutsche Lehrlingsausbildung? Mimeo.

Windmeijer, F.A.G. (1995): A Note on R^2 in the Instrumental Variables Model, Journal of Quantitative Economics 11, 257-261.

Wooldridge, J.M. (1995): Selection Correction for Panel Data Models under Conditional Mean Independence Assumptions, Journal of Econometrics 68, 115-132.

Wooldridge, J.M. (1999): Distribution-Free Estimation of Some Nonlinear Panel Data Models, Journal of Econometrics 90, 77-97.

Wooldridge, J.M. (2001): Econometric Analysis of Cross Section and Panel Data, MIT Press: Cambridge (Mass.).

Wright,P.G. (1928): Appendix to The Tariff on Animal and Vegatable Oils by P.G. Wright, Mcmillan: New York.

Yang, S.-S., Yu, Q. und M.A. Al-Zaid (2001): A Two Stage Estimation Procedure for Linear Mixed-Effects Models, Communications in Statistics, Theory and Methods 30, 2637-2653.

You, J. und G. Chen (2002): Parameter Estimation in a Partly Linear Regression Model with Random Coefficient Autoregressive Errors, Communication in Statistics - Theory and Methods 31, 1137 - 1158.

Zabel, J.E. (1992): Estimating Fixed and Random Effects with Selectivity, Economics Letters 40, 269-272.

Zheng, J.X. (1996): A Consistent Test of Functional Form via Non-parametric Estimation Function, Journal of Econometrics 75, 263 - 290.

Ziegler, A. und A. Eymann (2001): Zur Simulated Maximum-Likelihood-Schätzung von Mehrperioden-Mehralternativen-Probitmodellen, Allgemeines Statistisches Archiv 85, 319-342.