

# Eine Übersicht über gängige Modelle der Panelökonometrie und ihre kausale Interpretation

Michael Lechner\*

*Schweizerisches Institut für Aussenwirtschaft und Angewandte Wirtschaftsforschung (SIAW) der  
Universität St. Gallen  
CEPR (London), IZA (Bonn), ZEW (Mannheim)*

Erster Entwurf: September 2001

Gedruckt am 29.11.2001

**Kommentare sind herzlich willkommen**

*Vorbereitet für die Tagung der Deutschen Statistischen Gesellschaft, Dortmund, September  
2001*

## **Korrespondenzadresse**

Prof. Dr. Michael Lechner  
Schweizerisches Institut für Aussenwirtschaft und angewandte Wirtschaftsforschung (SIAW)  
Universität St. Gallen  
Dufourstr. 48, CH-9000 St. Gallen, Schweiz  
Michael.Lechner@unisg.ch,  
[www.siaaw.unisg.ch/lechner](http://www.siaaw.unisg.ch/lechner)

---

\* Ich bedanke mich beim Schweizerischen Nationalfonds (Projekte 4045-058311, 4045-050673) für die finanzielle Unterstützung. Ich danke ausserdem Reinhard Hujer und Markus Frölich, sowie den Teilnehmern der Tagung der Deutschen Statistischen Gesellschaft in Dortmund, 2001, für anregende Hinweise und Kommentare. Alle verbleibenden Fehler und Ungenauigkeiten sind meine eigenen.

## Zusammenfassung

Paneldaten gewinnen als Datenmaterial für empirische Studien im Bereich der Wirtschaftswissenschaften, sowie auch in vielen anderen empirisch ausgerichteten Forschungsfeldern, eine immer grössere Bedeutung. Das Papier gibt einen nichttechnischen Überblick über in der ökonometrischen Paneldatenanalyse populäre lineare und nichtlineare Modelle und diskutiert die Annahmen, unter denen die geschätzten Modellparameter eine kausale Interpretation besitzen.

### 1 Einleitung

Paneldaten gewinnen als Datenmaterial für empirische Studien im Bereich der Wirtschaftswissenschaften, sowie auch in vielen anderen empirisch ausgerichteten Forschungsfeldern eine immer grössere Bedeutung. *Paneldaten* zeichnen sich dadurch aus, dass für eine Einheit (Person, Firma, Land, etc.) Informationen für mindestens eine weitere Dimension (typischerweise Zeit oder Raum<sup>1</sup>) zur Verfügung stehen.

In den letzten Jahren war die Panelökonometrie sowohl im Bereich der Methodenentwicklung als auch im Bereich der anwendungsbezogenen Forschung ein ausserordentlich aktives Teilgebiet der Ökonometrie. Die Panelökonometrie hat (inzwischen) wiederum selbst zwei grössere Teilbereiche entwickelt. Zum einen gibt es die Literatur, die sich auf Paneldaten mit einer grossen Anzahl von Einheiten konzentriert. Diese Einheiten können wiederholt beobachtet werden. Der Stand der Forschung der in diesem zeitlich früher beginnenden Strang der panelökonometrischen Literatur inzwischen erreicht wurde, kann im wesentlichen in den beiden Beiträgen im *Handbook of Econometrics* von Gary Chamberlain (1984) und Manuel

---

<sup>1</sup> Ein Datensatz, der z.B. die Aktivitäten verschiedener Firmen in unterschiedlichen Regionen enthält, besitzt eine ganz ähnliche Struktur zu den Paneldaten, die Zeit als die zusätzliche Dimension besitzen (siehe beispielsweise Anselin und Bera, 1997). Jedoch wird sich dieser Aufsatz aus Platzgründen ausschliesslich auf Paneldatensätze mit den gängigeren Dimensionen *Einheit* und *Zeit* beschränken.

Arellano und Bo Honoré (2001) nachgelesen werden.<sup>2</sup> Man findet hier grosse Ähnlichkeiten zur klassischen Mikroökonomie: Die Daten sind meistens Individualdaten und es werden die durch Zensierung, Selektivität, etc., entstehenden Probleme explizit betrachtet. Bezüglich der Art der Asymptotik, die zur Approximation der stochastischen Eigenschaften der Schätzverfahren in endlichen Stichproben verwendet wird, aber auch bezüglich der Modellierung geht man von einer grossen (wachsenden) Anzahl (Querschnitts-) Einheiten und einer eher kleinen (konstanten) Anzahl von Zeitperioden aus. Dieser Aufsatz steht ebenfalls in dieser Tradition.<sup>3</sup>

In jüngerer Zeit wurden aber auch den Zeitreiheneigenschaften von Paneldaten vermehrte Aufmerksamkeit gewidmet (siehe z.B. Philips und Moon, 1999, oder Baltagi und Kao, 2001). Bezüglich der asymptotischen Inferenz geht man in diesen Studien oft, aber nicht immer, von einer festen Anzahl Querschnittsbeobachtungen und einer wachsenden Zeitdimension aus und verwendet auch eine darauf abgestimmte Modellierung.

Gleichzeitig zu den methodischen Aktivitäten wurden von verschiedenen Seiten vermehrt finanzielle Mittel eingesetzt, um der Wissenschaft umfangreiche Paneldatensätze zur Analyse zur Verfügung zu stellen. Beispielhaft sei der Bereich der angewandten Arbeitsmarkt- und Sozialforschung in Deutschland angeführt: 1984 begann mit dem Sozio-ökonomischen Panel (GSEOP) ein umfragebasierter Individualpaneldatensatz mit mehreren tausend Haushalten, der inzwischen seit 17 Jahren jedes Jahr um ein weiteres Jahr verlängert wird und sich bei empirisch arbeitenden Wirtschafts- und Sozialforschern grosser Beliebtheit erfreut. Ein Problem von umfragebasierten Datensätzen ist deren langsames *Wachstum* über die Zeit. Will man jedoch für einen Paneldatensatz von Beginn an eine grössere Zeitdimension erreichen, so

---

<sup>2</sup> Andere umfassende, auch Anwendungen einschliessende, Betrachtungen können zum Beispiel im 'Handbuch der Panelökonomie' von Mátyás und Sevestre (1996) gefunden werden.

<sup>3</sup> Es ist innerhalb der hier geltenden Seitenbeschränkungen natürlich unmöglich alle Facetten dieser reichhaltigen Literatur auch nur kurz darzustellen. Die grösste *Unterlassungssünde* betrifft wohl den Bereich der Verweildauermodelle, der hier vollkommen ausgeklammert wird (siehe z.B. die Übersichtsarbeit von van den Berg, 2001, für einen umfassenden Überblick über diesen Bereich). Auch wird die inzwischen stark angewachsene Literatur mit vielfältigen empirischen Anwendungen vollkommen ausgeklammert.

wartet man nicht auf künftige Umfragen, sondern greift auf schon anderweitig gespeicherte Informationen zurück. Ein Beispiel hierfür ist das sogenannte *public use file* der IAB-Beschäftigtenstichprobe (IAB-S), die aus administrativen Daten erzeugte Arbeitsmarktinformationen über eine grosse Anzahl von Personen für die letzten 25 Jahre enthält.<sup>4</sup> In der Schweiz entsteht zur Zeit ein Paneldatensatz, der durch eine Kombination von Umfragedaten und administrativen Daten zumindest für eine Reihe von Variablen eine (lange) Paneldimension aufweist.<sup>5</sup> Ähnliche Entwicklungen bezüglich der Zunahme von qualitativ hochwertigen Datensätzen mit Panelcharakter kann man in vielen anderen Anwendungsgebieten, wie zum Beispiel der Industrieökonomik und der Finanzmarktanalyse, und auch in anderen Ländern beobachten.

Was ist nun die Magie von Paneldaten, die dazu führt, so viel Geld und intellektuelles Kapital in diesen Bereich zu investieren? Warum werden Paneldaten als so viel nützlicher angesehen als reine Querschnittsdaten, die es erlauben sollten, individuelle Unterschiede zu beschreiben? Ähnliches gilt für reine Zeitreihen, die zur Beschreibung von dynamischen Phänomenen gut geeignet sein sollten? Von der methodischen und der Anwendungsseite betrachtet, scheint die Antwort einfach zu sein: Das Vorhandensein eines zweiten Querschnitts lässt die Annahme, dass viele Facetten der Struktur des datenerzeugenden Prozesses von einem Querschnitt zum anderen konstant sind, oft plausibel erscheinen. Genau das gleiche Argument lässt sich bei der Betrachtung mehrerer Zeitreihen über die gleichen Zeitpunkte hinweg anbringen, denn auch hier kann oft plausibel angenommen werden, dass viele dynamische Eigenschaften des datenerzeugenden Prozesses sich wiederholen. Mit anderen Worten, Paneldaten (kombiniert mit plausiblen Restriktionen über Zeit oder Einheiten hinweg) führen zu einer erhöhten

---

<sup>4</sup> Siehe auch die Übersicht in Hujer, Rendtl und Wagner (1997).

<sup>5</sup> Ab der Welle 2001 werden der Schweizerische Arbeitskräfte- und Erwerbsstichprobe (SAKE) historische Einkommensinformationen aus dem System der Arbeitslosen- und Hinterbliebenenversicherung (AHV) zugespielt (siehe Buhman, 2001). In der SAKE werden die gleichen Haushalte 4 Jahre lang hintereinander befragt. Jedes Jahr wird ein Viertel der Stichprobe durch neue Haushalte ersetzt. Es handelt sich also um ein sogenanntes rotierendes Panel. Die Qualität dieser retrospektiven Information sollte wesentlich höher sein, als wenn man die Personen in einer jeweiligen Umfrage nach lange zurückliegenden Informationen befragt.

Anzahl von Freiheitsgraden in der Schätzung. Diese zusätzlichen Freiheitsgrade können immer zur Effizienzsteigerung der Schätzverfahren verwendet werden. Oft kann man jedoch diese zusätzlichen Freiheitsgrade wieder 'aufbrauchen', indem das Modell verallgemeinert und somit zusätzliche Effekte identifiziert (und konsistent geschätzt) werden, die mit einem reinen Querschnitt oder einer reinen Zeitreihe allein nicht identifizierbar gewesen wären. So lässt sich z.B. dynamisches Verhalten analysieren, das von individueller Heterogenität betroffen ist.

Inzwischen hat sich die Panelökonometrie zu einem ausgereiften Forschungsgebiet entwickelt. So gibt es neben den schon erwähnten Übersichtsartikeln auch einige hervorragende Lehrbücher und Monographien, die das Feld weitgehend abdecken.<sup>6</sup> Auch ist das formale Niveau dieser Übersichtsarbeiten hinreichend heterogen, so dass eigentlich für jeden 'Geschmack' etwas angeboten wird. Was kann also der *value added* des hier angestrebten Übersichtsartikels sein? Betrachtet man die beiden Gebiete, die Paneldaten attraktiv machen, nämlich zusätzliche Identifikation und Effizienzgewinn, so scheint es aus dieser Sichtweise wenig sinnvoll, mit diesem Beitrag zum Beispiel das Thema Effizienzgewinn explizit aufzugreifen zu wollen, da den erwähnten Papieren und Büchern wenig hinzuzufügen wäre. Hinzu kommt, dass aufgrund der inzwischen oft anzutreffenden Stichprobengrößen in praktischen Anwendungen mögliche Effizienzgewinne eher von rein theoretischem Interesse sind. Identifikation bezieht sich jedoch auf die Möglichkeit, bestimmte Eigenschaften des datenerzeugenden Prozesses aus den Daten lernen zu *können* (wenn eine sehr grosse Stichprobe vorhanden wäre) und ist somit unabhängig von der Stichprobengröße.

In der Literatur über Paneldatenmodelle wird das Thema Identifikation oft aus der Sicht von spezifischen vorgegebenen Modellparametern betrachtet. Dieses Vorgehen hat in vielen Fällen, in denen das ökonometrische Modelle klar einen *kausalen* Zusammenhang beschreibt,

seine Berechtigung. Hier wird einen Schritt weitergegangen und die Frage gestellt unter welchen Umständen, das heisst Annahmen über den datenerzeugenden Prozess, die Parameter oft verwendeter Modelle, sofern sie überhaupt identifiziert sind, eine kausale Bedeutung haben. Es ist somit das Ziel dieses Papiers einen gewissen Überblick über übliche Paneldatenmodelle zu geben und gleichzeitig den Zusammenhang der zugrundeliegenden Annahmen zur Kausalitätsanalyse zu ziehen. Um einen möglichst einfachen analytischen Rahmen zu erhalten, wird nur der einfachste Fall eines kausalen Effektes einer binären Variablen auf die abhängige Variable der 'Schätzgleichung' untersucht.

Im nächsten Abschnitt wird die notwendige Notation eingeführt und der kausale Analyse-rahmen definiert. Abschnitt 3 diskutiert gängige lineare und nichtlineare Modelle mit zufälligen Effekten. Abschnitt 4 führt ähnliche Betrachtungen für Modelle mit sogenannten fixen Effekten durch. Der letzte Abschnitt zieht Schlussfolgerungen.<sup>7</sup>

## 2 Der Modellrahmen für die Kausalanalyse

In dieser Arbeit wird in klassischer mikroökonomischer Tradition von einem Paneldatensatz mit vielen Querschnittseinheiten ( $i = 1, \dots, N$ ) ausgegangen, die über eine begrenzte Anzahl von Perioden  $t = 1, \dots, T$  beobachtet werden können. Somit kann der datenerzeugende Prozess (in der Population) mit zeitlich indizierten Zufallsvariablen beschrieben werden. Für die beobachtbare Stichprobe der Grösse  $NT$  wird folgerichtig angenommen, dass sie aus  $N$  unabhängigen Ziehungen (aus der Population der Einheiten) besteht. Die Stichprobe ist somit in der Querschnittsdimension identisch und unabhängig verteilt, weist jedoch für die gleiche Querschnittseinheit Abhängigkeiten in der Längsschnittsdimension auf. Den üblichen

---

<sup>6</sup> Siehe zusätzlich zu den schon genannten Arbeiten beispielsweise die Monographie von Hsiao (1986), die Übersichtsarbeiten von Greene (2001) und Krishnakumar (1999), und die Lehrbücher von Arellano (2002), Baltagi (1995) und Lee (2002).

<sup>7</sup> Der kundige Leser wird schnell feststellen, dass die Papiere des Autors überproportional oft zitiert werden. Das liegt sicher nicht daran, dass diese so überragend wären, sondern vielmehr daran, dass ich diese am besten kenne!

Gepflogenheiten folgend, werden die abhängigen Variablen des Zeitpunkts  $t$  mit  $Y_t$  und die unabhängigen Variablen mit  $X_t$  bezeichnet. Es sind jeweils die Realisationen  $y_{it}$  und  $x_{it}$  beobachtbar.<sup>8</sup> Ausserdem gibt es wie üblich einen Fehlerterm  $U_t$ , von dem keine Realisationen beobachtet werden können. Das primäre Interesse des angewandten Forschers gelte einer besonderen Variablen  $D$ , die nur die Ausprägungen 0 und 1 annehmen kann. Ziel der Analyse ist es den kausalen Effekt dieser Variable  $D$  auf die abhängige Variable herauszufinden.<sup>9</sup>  $X_t$  und  $U_t$  sind nur in so weit von Interesse, wie sie behilflich sein können diesen kausalen Effekt aufzudecken (z.B. in dem man für ihren Einfluss auf die abhängige Variable kontrolliert bzw. diesen Einfluss durch Annahmen über ihre Verteilung einschränkt).

Unter welchen Umständen und mit welchen Kombinationen von Daten und (plausiblen) Annahmen bzw. Modellen lassen sich geschätzte Zusammenhänge, wie z.B. Regressionskoeffizienten, kausal interpretieren? Diese Frage beschäftigt seit langem nicht nur die Statistik und die Ökonometrie, sondern auch die Biometrie, die Epidemiologie, und im Grunde alle anderen Felder (auch wenn es nicht immer explizit erwähnt wird), denn es war und ist schon immer ein zentrales Anliegen der Wissenschaft, kausale Wirkungszusammenhänge aufzudecken und zu quantifizieren.

Die bisher eingeführte Notation ist für eine kausale Analyse jedoch noch nicht ausreichend. Inzwischen hat sich in der Ökonometrie und auch zumindest in Teilen der Statistik das 1974 in einem Aufsatz von Donald Rubin formalisierte kausale Modell potentieller Ergebnisse als geeigneter Analyserahmen durchgesetzt. Dieses Modell beschreibt den kausalen Effekt als eine Differenz von potentiellen Ergebnissen, die sich unter verschiedenen Werten der interessierenden kausalen Variablen ( $D$ ) einstellen *würden*. Das potentielle Ergebnis, das sich in

---

<sup>8</sup> Grossbuchstaben bezeichnen Zufallsvariablen, kleine Buchstaben bezeichnen entweder Realisationen oder bestimmte Werte dieser Zufallsvariablen.

<sup>9</sup> Die Annahme die Variable  $D$  könne nur zwei Werte annehmen wird lediglich zur Vereinfachung der Notation und Argumentation getroffen. Der allgemeinere Fall wird ausführlich in Imbens (2000) und Lechner (2001 a, b, c) diskutiert, allerdings ohne den Bezug zu Paneldatenmodellen herzustellen.

Periode  $t$  für  $D=1$  ergibt, wird als  $Y_t^1$  bezeichnet, während das korrespondierende potentielle Ergebnis für  $D=0$  mit  $Y_t^0$  bezeichnet wird.<sup>10</sup> Der mittlere kausale Effekt für den Teil der Population mit Charakteristika  $x_t$  ist wie folgt definiert:

$$\gamma_{t,0}(x_t) = E(Y_t^1 | X_t = x_t) - E(Y_t^0 | X_t = x_t), \quad \forall x_t, \quad \forall t. \quad (1)$$

Da sich jede Einheit zu jedem Zeitpunkt nur in einem Zustand befindet, können die beiden potentiellen Ergebnisse für eine Einheit zum selben Zeitpunkt nicht gleichzeitig beobachtet werden. Daher erhält man für die Beziehung zwischen den beobachtbaren abhängigen Variablen und den potentiellen abhängigen Variablen folgende Beziehung (die sogenannte Beobachtungsregel):

$$Y_t = (1 - D)Y_t^0 + DY_t^1. \quad (2)$$

Zusätzlich soll angenommen werden, dass die Variable  $D$  erst ab Periode  $s$  ( $1 < s \leq T$ ) eine Wirkung entfaltet (z.B. wenn es sich bei  $D$  um ein unerwartetes Ereignis handelt, das in Periode  $s$  eingetreten ist). Daher gilt für alle Perioden vor Periode  $s$  folgende Beziehung:

$$Y_t = Y_t^0 = Y_t^1 \quad \Rightarrow \quad \gamma_{t,0} = 0 \quad \forall t = 1, \dots, s-1. \quad (3)$$

Zur Vervollständigung der Notation ist es nützlich, eine zeitlich indizierte Variable  $D_t$  einzuführen, die sich wie folgt aus Annahme (3) ergibt:

$$D_t = \underline{1}(t \geq s) D. \quad (4)$$

---

<sup>10</sup> Die für dieses Modell notwendigen weiteren Annahmen, die in Literatur als SUTVA und Common Support bezeichnet werden, sind für diese Ausführungen nicht zentral und werden als erfüllt angesehen (siehe Rubin, 1974).



Die Indikatorfunktion  $\underline{1}(\cdot)$  sei eins, wenn die Bedingung in der Klammer wahr ist, und null sonst. Bezeichnet man die Realisation der Zufallsvariablen  $D_t$  als  $d_{it}$  ( $d_{it}=1$  gleich eins, wenn die Beobachtung zu einem Zeitpunkt  $s$  oder später gemessen wurde und die Einheit  $i$  zur Gruppe der Personen mit  $d=1$  gehört), so können die Daten wie folgt beschrieben werden:

$$\{y_{it}, d_{it}, x_{it}\}, \quad t=1, \dots, T; \quad i=1, \dots, N. \quad (5)$$

Ohne weitere Annahmen lässt sich allerdings aus den schätzbaren Momenten der beobachtbaren Variablen  $Y$ ,  $D$  und  $X$  offensichtlich nicht auf den in Gleichung (1) definierten kausalen Effekt schliessen. Dies sieht man direkt, wenn man den in Gleichung (1) definierten Effekt in Termini der beobachtbaren Variablen (Gleichung (2)) schreibt:

$$\begin{aligned} \gamma_{t,0}(x_t) &= E(Y_t^1 | X_t = x_t) - E(Y_t^0 | X_t = x_t) \\ &= [E(Y_t^1 | D_t = 1, \cdot) - E(Y_t^0 | D_t = 1, \cdot)]P(D_t = 1 | X_t = x_t) + [E(Y_t^1 | D_t = 0, \cdot) - E(Y_t^0 | D_t = 0, \cdot)][1 - P(D_t = 1 | \cdot)] \\ &= [E(Y_t | D_t = 1, \cdot) - E(Y_t^0 | D_t = 1, \cdot)]P(D_t = 1 | \cdot) + [E(Y_t^1 | D_t = 0, \cdot) - E(Y_t | D_t = 0, \cdot)][1 - P(D_t = 1 | \cdot)]; \\ &\quad \forall t \geq s. \quad (6) \end{aligned}$$

Während es sich bei  $E(Y_t | D_t = 1, \cdot)$  und  $E(Y_t | D_t = 0, \cdot)$  um z.B. mit Hilfe von korrespondierenden Stichprobenmomenten schätzbare Grössen handelt, lassen sich die Ausdrücke  $E(Y_t^0 | D_t = 1, \cdot)$  und  $E(Y_t^1 | D_t = 0, \cdot)$  nicht ohne weitere Annahmen in Ausdrücke umformulieren, von denen Stichprobenäquivalente beobachtbar wären. Eine der mögliche Annahmen wäre, dass die Zufallsvariable  $D_t$  ihre Werte unabhängig von den Ergebnisvariablen annimmt (wie z.B. bei Experimenten mit entsprechender Randomisierung), denn dann würde  $E(Y_t^0 | D_t = 1, \cdot) = E(Y_t^0 | D_t = 0, \cdot) = E(Y_t | D_t = 0, \cdot)$  und  $E(Y_t^1 | D_t = 1, \cdot) = E(Y_t | D_t = 1, \cdot)$ .

Den Unterschied zwischen dem *counterfactual*  $E(Y_t^0 | D_t = 1, \cdot)$  und dem schätzbaren Term  $E(Y_t^0 | D_t = 0, \cdot)$  bezeichnet man auch als Selektionsverzerrung, da er daher kommt, dass Einheiten mit  $d = 0$  andere Erwartungswerte von  $Y_t^0$  besitzen, als Einheiten für die  $d = 1$  beobachtet werden kann.

Für die Ausführung in diesem Papier spielen Experimente keine Rolle, jedoch spielen die folgenden beiden einem experimentellen Kontext sehr ähnlichen Annahmen eine Rolle, die man als Selektion auf Basis beobachtbarer Variablen (*selection on observables*) und Selektion auf Basis (teilweise) unbeobachtbarer Variablen (*selection on unobservables*) bezeichnen kann. Bei der Selektion auf Basis beobachtbarer Variablen geht man davon aus, dass man mit dem vorhandenen Datensatz für alle Variablen ( $X$ ) kontrollieren kann, die gleichzeitig die potentiellen Ergebnisse und die interessierenden Variablen beeinflussen. In diesem Falle müssen für gegebene Werte dieser Kontrollvariablen die potentiellen Ergebnisse und  $D$  unabhängig sein. Somit lassen sich  $E(Y_t^1 | X_t = x_t)$  und  $E(Y_t^0 | X_t = x_t)$  wie folgt als Erwartungswerte beobachtbarer Größen ausdrücken:<sup>11</sup>

$$E(Y_t^1 | X_t = x_t) = E(Y_t^1 | D_t = 1, X_t = x_t) = E(Y_t | D_t = 1, X_t = x_t), \quad (7)$$

$$E(Y_t^0 | X_t = x_t) = E(Y_t^0 | D_t = 0, X_t = x_t) = E(Y_t | D_t = 0, X_t = x_t). \quad (8)$$

Verwendet man auch noch zusätzlich die Annahme homogener Effekte so lässt sich der kausale Effekt unter der Annahme homogener Effekte wie definiert in Gleichung (1) folgendermassen umschreiben:

$$\gamma_{t,0}(x_t) = E(Y_t^1 | X_t = x_t) - E(Y_t^0 | X_t = x_t) = E(Y_t | D_t = 1, X_t = x_t) - E(Y_t | D_t = 0, X_t = x_t). \quad (9)$$

---

<sup>11</sup> Diese Annahme wird in der einschlägigen Literatur auch als *conditional independence assumption* (CIA) bezeichnet.

Gilt die den Gleichungen (7) und (8) zugrundeliegende Annahme nicht, muss es weitere nicht beobachtbare Variablen geben, die sowohl  $D$  als auch die potentiellen Ergebnisse beeinflussen. Man spricht dann von Selektion auf Basis zumindest teilweise unbeobachtbarer Variablen. Es lässt sich in diesem Fall eine zu den Gleichungen (7) und (8) ähnliche Gleichung definieren, indem man eine unbeobachtbare Zufallsvariable genau derart definiert, dass folgende Gleichungen gültig sind:

$$E(Y_t^1 | X_t = x_t) = E_{V_t|X_t=x_t} E(Y_t^1 | X_t = x_t, V_t) = E_{V_t|X_t=x_t} E(Y_t^1 | D_t = 1, X_t = x_t, V_t) = E_{V_t|X_t=x_t} E(Y_t | D_t = 1, X_t = x_t, V_t). \quad (10)$$

$$E(Y_t^0 | X_t = x_t) = E_{V_t|X_t=x_t} E(Y_t^0 | X_t = x_t, V_t) = E_{V_t|X_t=x_t} E(Y_t^0 | D_t = 0, X_t = x_t, V_t) = E_{V_t|X_t=x_t} E(Y_t | D_t = 0, X_t = x_t, V_t). \quad (11)$$

Im Gegensatz zu den Gleichungen (7) und (8) sind die Gleichungen (10) und (11) nicht direkt für die empirische Analyse nutzbar, da  $V_t$  unbeobachtbar ist. Später wird jedoch gezeigt werden, dass es unter gewissen weiteren Annahmen über die unbekannte Variable  $V_t$  mit Paneldaten möglich ist, kausale Effekte zu identifizieren, die mit Querschnittsdaten nicht identifizierbar wären, gerade wenn es sich um den schwierigen Fall von Selektion auf Basis unbeobachtbarer Variablen handelt. Unter Verwendung von Gleichungen (10) und (11) kann man den Effekt in diesem Fall auch wie folgt schreiben:

$$\gamma_{t,0}(x_t) = E_{V_t|X_t=x_t} E(Y_t | D_t = 1, X_t = x_t, V_t) - E_{V_t|X_t=x_t} E(Y_t | D_t = 0, X_t = x_t, V_t). \quad (12)$$

In den nächsten beiden Abschnitten werden häufig für die ökonometrische Panelanalyse verwendete Modelle vorgestellt, und daraufhin untersucht, welche der hier aufgeführten Annahmen zur kausalen Interpretation der geschätzten (Koeffizienten-) Werte mit dem jeweiligen Modell kompatibel sind.

### 3 Modelle mit zufälligen Effekten

#### 3.1 Allgemeine Bemerkungen

Die Klasse der parametrischen Modelle mit zufälligen Effekten (*random effects models*) zeichnet sich durch zwei wesentliche Annahmen aus: Zum einen ist der bedingte Erwartungswert der abhängigen Variablen gegeben die erklärenden Variablen eine bekannte Funktion. Zum anderen sind die als Differenz der Zufallsvariablen (ZV) und ihrem konditionalen Erwartungswert definierten Residuen über die Zeit korreliert. Oft wird zusätzlich angenommen, dass diese Korrelationen zeitkonstant sind.

Die Panelstruktur dient in dieser Modellklasse neben dem Aufdecken der dynamischen Veränderungen der konditionalen Momentenfunktion und der Dynamik der Fehlerterme im wesentlichen dazu, die Effizienz der Schätzung durch das Hinzuziehen zusätzlicher Beobachtungen von anderen Zeitpunkten zu steigern.<sup>12</sup> Die folgenden Erläuterungen werden zeigen, dass die typischen Schätzverfahren des *random effects* Modells unter zwei Annahmen eine kausale Interpretation der geschätzten Koeffizienten erlauben: Zum einen muss der bedingte Erwartungswert der beobachtbaren abhängigen Variablen korrekt spezifiziert sein, und zum anderen darf es sich nur um ein Selektionsproblem auf Basis von beobachtbaren Variablen handeln.

#### 3.2 Das lineare Modell

Das klassische lineare Panelregressionsmodell ist durch den folgenden datengenerierenden Prozess beschrieben:

$$E(Y_t | D_t = d_t, X_t = x_t) = d_t \delta_0 + x_t \beta_0 \quad \rightarrow \quad Y_t = D_t \delta_0 + X_t \beta_0 + U_t \quad \text{mit} \quad E(U_t | D_t, X_t) = 0. \quad (13)$$

---

<sup>12</sup> In der Tat wäre jedoch die Effizienzsteigerung in vielen Fällen noch grösser, hätte man statt einem Panel wiederholte Querschnitte (also in jede Welle neue Einheiten) zur Verfügung, bei denen die einzelnen Beobachtungen nicht (konditional positiv) korreliert sind.

Die Kovarianzmatrix des Fehlerterms kann entweder unbeschränkt bleiben ( $E(U_t U_\tau | D_t, X_t, D_\tau, X_\tau) = \sigma_{t\tau}$ ), oder man kann Äquikorrelation auferlegen, indem man die populäre Fehlertermzerlegung mit einem zeitkonstanten Individualeffekt ( $C$ ), der mit den Regressoren und dem verbleibenden Fehlerterm unkorreliert ist, unterstellt:<sup>13</sup>

$$U_t = C + E_t \quad \text{mit} \quad E(C | \cdot) = 0, \quad E(E_t | \cdot) = 0, \quad E(CE_t | \cdot) = 0, \quad E(E_t E_\tau | \cdot) = 0,$$

$$\text{Var}(C | \cdot) = \sigma_c^2, \quad \text{Var}(E_t | \cdot) = \sigma_E^2, \quad \forall t, \tau \quad t \neq \tau. \quad (14)$$

Daraus folgt direkt die Äquikorrelation (d.h. die Korrelation zweier Beobachtungen der gleichen Einheit ist unabhängig von der betrachteten zeitlichen Distanz), denn es gilt:

$$\text{Cov}(U_t, U_\tau | \cdot) = E(C^2 + \underbrace{CE_t}_0 + \underbrace{CE_\tau}_0 + \underbrace{E_t E_\tau}_0 | \cdot) = \sigma_c^2 \rightarrow \text{Corr}(U_t, U_\tau | \cdot) = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_E^2}. \quad (15)$$

Sind die üblichen Annahmen über die Regressoren erfüllt, so ist offensichtlich, dass die Standardannahmen des linearen Regressionsmodells gelten (unabhängig von der wahren intertemporalen Kovarianzstruktur der Fehlerterme) und die Koeffizienten  $\delta_0$  und  $\beta_0$  mit Hilfe einer einzigen Welle ( $t \geq s$ <sup>14</sup>) mit OLS konsistent geschätzt werden können.<sup>15</sup> Mit anderen Worten, die folgende bedingte Momentenrestriktion ist ausreichend zur Identifikation der Parameter des Modells:

$$E(Y_t - d_t \delta_0 - x_t \beta_0 | D_t = d_t, X_t = x_t) = 0. \quad \forall t, \quad t \geq s. \quad (16)$$

<sup>13</sup> In diesem Papier wird immer von strikter Exogenität ausgegangen, ohne dies immer explizit in den Konditionierungsbedingungen zu berücksichtigen. Das heisst, wenn eine Eigenschaft gegeben den Wert der Variablen der Periode angeben wird, wird immer davon ausgegangen, dass ein zusätzliches Konditionieren auf die gleichen Variablen aller anderen zukünftigen und vergangenen Perioden nichts an der jeweiligen Eigenschaft ändert. Der Beitrag von Richard Blundell wird explizit diese Annahme aufgeben und Implikationen diskutieren.

<sup>14</sup> Für  $t < s$  tritt ein Multikollinearitätsproblem auf, da in diesem Fall immer  $d_{it} = 0$  gilt.

Weitere Wellen können hinzugenommen (gepoolt) werden, und OLS (mit für die spezifische Korrelation der Fehlerterme korrigierten Standardfehlern) kann auf dem gepoolten Datensatz geschätzt werden (insbesondere erlauben es weitere Wellen die Koeffizienten  $\delta_0$  und  $\beta_0$  als zeitabhängig zu betrachten und die Kovarianzstruktur der Fehlerterme zu schätzen; dies soll jedoch hier nicht thematisiert werden). Asymptotisch effizienter ist eine FGLS Schätzung, wobei die OLS-Residuen verwendet werden können, um die GLS Gewichtungsmatrix konsistent zu schätzen. Diskussionen über die Eigenschaften dieser Schätzverfahren, oder anderer in Frage kommender Schätzverfahren werden in jedem der erwähnten Lehrbücher zur Panelökonometrie ausführlich diskutiert.

Nun stellt sich die Frage, inwieweit die Momentenbedingungen auf Basis der beobachtbaren Variablen den kausalen Effekte identifizieren können. Mit anderen Worten, unter welchen Annahmen kann der Koeffizient  $\delta_0$  (bzw. dessen konsistente Schätzung  $\hat{\delta}_N$ ) als kausaler Effekt ( $\gamma_0$ ) interpretiert werden? Die wesentliche Restriktion des linearen Modells lautet  $E(Y_t | D_t = d_t, X_t = x_t) = d_t \delta_0 + x_t \beta_0$ . Somit entspricht  $\delta_0$  genau der Differenz  $E(Y_t | D_t = 1, X_t = x_t) - E(Y_t | D_t = 0, X_t = x_t)$ . Diese Differenz entspricht dem kausalen Effekt im Falle einer Selektion auf beobachtbaren Variablen (siehe Gleichung (9)), denn nur dann gilt  $E(Y_t | D_t = 1, X_t = x_t) = E(Y_t^1 | X_t = x_t)$  und  $E(Y_t | D_t = 0, X_t = x_t) = E(Y_t^0 | X_t = x_t)$ .

Zusammengefasst lässt sich folgendes festhalten: Falls es sich um eine Selektion auf Basis beobachtbarer Variablen handelt, die konditionalen Erwartungswerte der potentiellen Ergebnisse folgende lineare Form besitzen,

$$E(Y_t^1 | X_t = x_t) = \delta_0 + x_t \beta_0 \quad \text{und} \quad E(Y_t^0 | X_t = x_t) = x_t \beta_0, \quad (17)$$

---

<sup>15</sup> Der Konsistenzbegriff wird in dieser Arbeit im allgemeinen in dem Sinne  $N \rightarrow \infty$  bei konstantem  $T$  verwendet.

besitzen und die kausalen Effekte nicht mit den Kontrollvariablen variieren, dann liefert das *random effects* Panelmodell eine konsistente Schätzung des kausalen Effekts von  $D$  auf  $Y$ . Hierfür ist nur eine einzige Welle zum Zeitpunkt  $t$  ( $t \geq s$ ) notwendig. Durch Einfügen von geeigneten Interaktionstermen kann das Modell bezüglich der Annahmen der Effekthomogenität und der funktionalen Form des Erwartungswertes (wie üblich) flexibilisiert werden, jedoch bleibt die grundsätzliche Restriktion, dass es sich nur um Selektion auf Basis von beobachtbaren Variablen handeln darf, bestehen.

### 3.3 Das nichtlineare Modell

In vielen Fällen wird die Linearitätsannahme der Art der abhängigen Variablen nicht gerecht. Wird der Erwartungswert nichtlinear modelliert, dann lässt sich der angenommene datenerzeugende Prozess in allgemeiner Form in dieser Modellklasse wie folgt schreiben:

$$Y_t = \underbrace{g_t(D_t, X_t; \beta_0, \delta_0)}_{\equiv E(Y_t | D_t, X_t)} + U_t \quad \text{mit } U_t \equiv Y_t - g_t(D_t, X_t; \beta_0, \delta_0) \rightarrow E(U_t | D_t, X_t) = 0, \quad \forall t. \quad (18)$$

Enthält Gleichung (18) genau die bedingten Momentenrestriktionen, die zur Identifikation der unbekannt Parameter  $\beta_0$ ,  $\delta_0$  notwendig sind, so führt jede weitere Periode (lediglich) zu einem Effizienzgewinn.<sup>16</sup> Bertschek und Lechner (1998) und Inkmann (2000) diskutieren die Struktur des asymptotisch effizienten Generalised Methods of Moments (GMM) Schätzverfahren in dieser Situation.<sup>17</sup>

Avery, Hansen und Hotz (1983), Guilkey und Murphy (1993) und Bertschek und Lechner (1998) diskutieren konkrete Schätzverfahren für das binäre Probitmodell, das im folgenden

<sup>16</sup> Es sei angenommen, dass die üblichen Anforderungen bzgl. Anzahl und Art der Regressoren und der Ableitungen der Funktion  $g_t(\cdot)$  erfüllt seien (keine Multikollinearität etc.).

<sup>17</sup> Lechner und Breitung (1996) diskutieren Schätzverfahren und Spezifikationstests in einer allgemeinen Klasse von nichtlinearen Modellen.

beispielhaft zur Beschreibung der nichtlinearen Modelle herangezogen werden soll. Das Probitmodell ist durch folgende latent lineare Struktur charakterisiert:<sup>18</sup>

$$Y_t = \mathbb{1}(D_t\delta_0 + X_t\beta_0 + U_t \geq 0), \quad \forall t, \quad \begin{bmatrix} U_1 \\ \vdots \\ U_T \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{U_1}^2 & \cdots & \sigma_{U_1 U_T} \\ & \ddots & \vdots \\ & & \sigma_{U_T}^2 \end{bmatrix} \right). \quad (19)$$

Somit ergibt sich für den Ausdruck des konditionalen Erwartungswertes der beobachtbaren binären abhängigen Variablen folgende Gleichung und die korrespondierende bedingte Momentenrestriktion:

$$E(Y_t | D_t = d_t, X_t = x_t) = \Phi\left(\frac{d_t\delta_0 + x_t\beta_0}{\sigma_{U_t}}\right), \quad \forall t; \quad (20)$$

$$E(Y_t - \Phi\left(\frac{D_t\delta_0 + X_t\beta_0}{\sigma_{U_t}}\right) | D_t = d_t, X_t = x_t) = 0. \quad (21)$$

Wie in Modellen der binären Wahl üblich, sind in diesem Panelmodell in jedem Querschnitt die skalierten Koeffizienten  $\delta_0/\sigma_{U_t}$  und  $\beta_0/\sigma_{U_t}$  identifiziert. Somit führt eine konsistente Schätzung der skalierten Koeffizienten zu einer konsistenten Schätzung der bedingten Erwartungswerte der abhängigen Variablen. Das Zusammenfassen (poolen) mehrerer Wellen erlaubt eine effizientere Schätzung der Koeffizienten, da mehr Momentenbedingungen wie in Gleichung (21) zur Verfügung stehen. Dieser Effizienzgewinn kommt insbesondere dann zum Tragen, wenn man die Gleichheit der Fehlertermvarianzen über die Zeit annehmen kann.

Die Koeffizienten im Probitmodell haben keine direkte kausale Interpretation, da sie nicht in direktem Zusammenhang zur Differenz zweier Erwartungswerte stehen. Ist man jedoch bereit,

---

<sup>18</sup> Der Fehlerterm im latenten linearen Modell ist offensichtlich anders definiert, als der Fehlerterm im beobachtbaren Modell definiert in Gleichung (18). Es hat sich in der Ökonometrie weitgehend eingebürgert



die Annahme einer Selektion von der Basis beobachtbarer Variablen zu treffen, und die folgenden Annahmen über die funktionale Form der konditionalen Erwartungswerte der potentiellen abhängigen Variablen zu treffen,

$$E(Y_t^1 | X_t = x_t) = \Phi\left(\frac{\delta_0 + x_t \beta_0}{\sigma_{U_t}}\right), \quad \text{und} \quad E(Y_t^0 | X_t = x_t) = \Phi\left(\frac{x_t \beta_0}{\sigma_{U_t}}\right), \quad \forall t, \quad (22)$$

kann man den kausalen Effekt auf Basis der geschätzten Koeffizienten und der Daten gewinnen. Somit ergibt sich für den bedingten kausalen Effekt folgender Ausdruck:

$$\gamma_{t,0}(x_t) \equiv E(Y_t^1 | X_t = x_t) - E(Y_t^0 | X_t = x_t) = \Phi\left(\frac{\delta_0 + x_t \beta_0}{\sigma_U}\right) - \Phi\left(\frac{x_t \beta_0}{\sigma_U}\right). \quad (23)$$

Damit wird zum einen deutlich, dass der kausale Effekt eine monoton zunehmende Funktion im Koeffizienten  $\delta_0$  ist und der Effekt mit den Werten der Kontrollvariablen auf bestimmte (restriktive) Art und Weise variiert.

Abschliessend lässt sich analog zum linearen Modell festhalten, dass wenn die angenommene Struktur des bedingten Erwartungswertes korrekt ist, es sich um ein Problem von Selektion auf beobachtbare Variablen handelt, und die Abhängigkeit des Effekts der durch die Probitannahmen implizierten Struktur folgen, dann sind zwar die Koeffizienten im nichtlinearen Modell zwar im Allgemeinen nicht direkt kausal interpretierbar (mit Ausnahme von  $\delta_0 = 0$ , denn dies entspricht  $\gamma_{t,0}(x_t) = 0$ ), jedoch lassen sich aus ihnen durch geeignete Umformungen und weitere Berechnungen die kausalen Effekte konsistent schätzen. Es verbleibt abschliessend anzumerken, dass die meisten anderen nichtlinearen Modelle eine ganz ähnliche Struktur aufweisen, auch wenn insbesondere die Art der Heterogenität der Effekte ganz spezifisch von der unterstellten funktionalen Form der bedingten Erwartungswerte abhängt.

---

Annahmen über den Fehlerterm im latenten Modell zu treffen und daraus die Eigenschaften der abhängigen

## 4 Modelle mit fixen Individualeffekten (*fixed effects*)

### 4.1 Allgemeine Bemerkungen

Eine der Attraktionen konventioneller Modelle für Paneldaten ist es, dass sie plausible Restriktionen über die Panelwellen anbieten, die ausgenutzt werden können, um den Einfluss unbeobachtbarer Variablen (modelliert als Komponenten des Fehlerterms), die mit den unabhängigen Variablen des Modells korreliert sind, auszuschalten. Diese Modelle implizieren somit Momentenbedingungen, die nicht von diesen Fehlerkomponenten abhängen. Diese Momentenbedingungen sind mehr oder weniger offensichtlich in den linearen Modellen, nicht jedoch in vielen im Fall mit zufälligen Effekten üblichen nichtlinearen Modellen.<sup>19</sup>

Bezüglich der Nützlichkeit bei Identifikation kausaler Effekte wird sich zeigen, dass in einer spezifischen Form nun die ‚Kontrolle‘ für Selektivität auf Basis unbeobachtbarer Variablen möglich ist, sofern deren Einfluss über die Zeit konstant bleibt.

### 4.2 Das lineare Modell

Dem linearen Regressionsmodell mit fixen Effekten liegen folgende Annahmen über den datengenerierenden Prozess zugrunde:

$$E(Y_t | D_t = d_t, X_t = x_t, C = c) = d_t \delta_0 + x_t \beta_0 + c \rightarrow E(Y_t - D_t \delta_0 - X_t \beta_0 - C | D_t = d_t, X_t = x_t, C = c) = 0$$

$$\rightarrow Y_t = D_t \delta_0 + X_t \beta_0 + \underbrace{C + E_t}_{U_t} \text{ mit } E(U_t | D_t, X_t) \neq 0, E(E_t | D_t, X_t) = E(E_t | D_t, X_t, C) = 0. \quad (24)$$

Das Modell enthält eine Fehlerkomponente, die mit den Regressoren korreliert sein darf. Die entscheidende Einschränkung ist allerdings, dass der Einfluss dieser Komponente zeitkonstant

---

Variablen abzuleiten.

<sup>19</sup> Bei der Interpretation der Begriffe *fixed* und *random effects* ist eine gewisse Verwirrung unter einigen Anwendern zu beobachten. Es geht substantiell überhaupt nicht darum, ob irgendwelche Effekte stochastisch

ist. Die in Gleichung (24) wiedergegebenen bedingten Momentenrestriktionen können nicht direkt verwendet werden, da sie unbeobachtbar sind.<sup>20</sup> Besitzt man jedoch eine weitere Zeitperiode  $\tau$  (und verwendet die Annahme der strikten Exogenität explizit), so kann man die Momentenbedingungen der beiden Zeitperioden derart kombinierten, dass sie nicht mehr von  $C$  abhängen:

$$E(Y_t | D_t = d_t, X_t = x_t, D_\tau = d_\tau, X_\tau = x_\tau, C = c) - E(Y_\tau | \cdot) = (d_t - d_\tau)\delta_0 + (x_t - x_\tau)\beta_0. \quad (25)$$

Drückt man Gleichung (25) in der üblichen Form der Residuen aus (und beachtet, dass die Konditionierung auf  $C$  nicht notwendig ist, da Gleichung (26) nicht von  $C$  abhängt), resultieren die folgenden bedingten Momentenrestriktionen:

$$E(\underbrace{U_t - U_\tau}_{=E_t - E_\tau} | D_t = d_t, X_t = x_t, D_\tau = d_\tau, X_\tau = x_\tau) = E[(Y_t - Y_\tau) - (D_t - D_\tau)\delta_0 - (X_t - X_\tau)\beta_0 | \cdot] = 0. \quad (27)$$

Diese Momentenbedingungen hängen nur von beobachtbaren Grössen ab und können daher zur Schätzung der Koeffizienten  $\delta_0$  und  $\beta_0$  verwendet werden. Übliche Schätzverfahren basieren auf einer Transformation der Daten, welche die Abhängigkeit von  $C$  beseitigt (Bildung erster Differenzen, Abziehen des individuellen Mittelwerts über die Zeit – die sogenannte within - Transformation, etc.) und anschliessender OLS oder GLS Schätzung auf Basis der transformierten Daten.

Da die Möglichkeit der *Kontrolle* für eine unbeobachtete Variable, die mit den beobachtbaren Kontrollvariablen korreliert ist, fast etwas Magisches hat, ist es sicherlich nicht verwunder-

---

oder deterministisch sind. Es geht ausschliesslich darum, ob der Fehlerterm eine Komponente enthält, die mit den Regressoren korreliert sein könnte (*fixed effects*), oder nicht (*random effects*).

<sup>20</sup> Eine Alternative wäre diese  $N$  über die Zeit konstanten Komponenten als unbekannte Koeffizienten zu behandeln und mitzuschätzen. Da die Anzahl dieser Koeffizienten jedoch mit  $N$  wächst, ist für eine konsistente Schätzung  $T \rightarrow \infty$  notwendig ( $N$ - $T$ -Asymptotik, die bei kleinem  $T$  allerdings eine schlechte Approximation der Stichprobenverteilung darstellt). Umgekehrt muss man sich bei einer  $N$ -Asymptotik um über die Individuen konstante Zeiteffekte keine 'Sorgen' machen, da diese als Koeffizienten konsistent mitgeschätzt werden können.

lich, dass dieses sehr mächtige Tool der *fixed effects* Schätzung oft zu blauäugig angewandt wurde. Daher soll auch hier explizit darauf hingewiesen werden, dass neben der Annahme eines in allen Komponenten linearen Erwartungswertes zusätzlich angenommen werden muss, der mittlere Einfluss der problematischen Variable auf  $Y$  sei zeitkonstant, linear und unkorreliert mit dem verbleibenden Störterm gegeben zukünftige, kontemporäre und vergangene Werte der abhängigen Variablen (strikte Exogenität).

Im Gegensatz zum Modell mit zufälligen Effekten kann  $\delta_0$  eine kausale Interpretation besitzen, selbst wenn  $D$  gegeben die Kontrollvariablen mit weiteren unbeobachtbaren Variablen korreliert ist, die auch die potentiellen Ergebnisse beeinflussen. Jedoch muss der Einfluss dieser Variablen auf das Ergebnis gewissen weiteren Restriktionen unterworfen sein.

Geht man von Selektivität auf Basis unbeobachtbarer Variablen, die beide potentiellen Ergebnisse in gleicher Weise und zeitkonstant beeinflussen, und von  $X$  unabhängigen Effekten aus, so sieht man an folgenden Gleichungen, dass  $\delta_0$  eine kausale Interpretation besitzt:

$$E(Y_t^1 | X_t = x_t, C = c) = \delta_0 + x_t \beta_0 + c \quad \text{und} \quad E(Y_t^0 | X_t = x_t, C = c) = x_t \beta_0 + c. \quad (28)$$

$$\rightarrow E(Y_t^1 | X_t = x_t, V_t = c) - E(Y_t^0 | X_t = x_t, V_t = c) = \delta_0 + x_t \beta_0 + c - x_t \beta_0 - c = \delta_0. \quad (29)$$

Ein wichtiger Hinweis für die praktische Anwendung dürfte sein, dass es notwendig ist, im Panel jeweils mindestens eine Periode vor ( $t < s$ ), und eine Periode nach ( $t \geq s$ ) dem Beginn der Wirkung von  $D$  zu beobachten. Ansonsten wäre der Regressor  $D_t$  zeitkonstant und liesse sich (wie alle anderen zeitkonstanten Variablen mit zeitkonstanten Koeffizienten) nicht von  $C$  unterscheiden. Mit anderen Worten,  $\delta_0$  ist überhaupt nur identifiziert, wenn  $D_t$  über die Zeit

variiert (alternativ könnte natürlich auch  $D_t$  zeitkonstant sein und  $\delta_0$  über die Zeit variieren).<sup>21</sup>

### 4.3 Das nichtlineare Modell

Betrifft die Nichtlinearität nur den Modellteil mit beobachtbaren Variablen und tritt die unbeobachtbare, problematische Variable weiterhin additiv auf, so ändert sich wenig, ausser dass auf nichtlineare Schätzverfahren zurückgegriffen werden muss.

$$E(Y_t | D_t = d_t, X_t = x_t, C = c) = g_t(d_t, x_t; \delta_0, \beta_0) + c$$

$$\rightarrow E(Y_t - Y_\tau - g_t(d_t, x_t; \delta_0, \beta_0) - g_\tau(d_\tau, x_\tau; \delta_0, \beta_0) | D_t = d_t, D_\tau = d_\tau, X_t = x_t, X_\tau = x_\tau). \quad (30)$$

Bezüglich der kausalen Effekte gelten die in Abschnitt 3.3 und 4.2 dargestellten Überlegungen bezüglich der Heterogenität des Effekts (3.3) und des Herausdifferenzierens der unbeobachtbaren Komponenten (4.2).

Interessanter ist der Fall, wenn die problematische Variable vom übrigen Erwartungswert nicht additiv separabel getrennt werden kann, denn dann funktioniert der einfache 'Differenzierungstrick' nicht mehr. Leider tritt diese Nichtseparabilität bei nahezu alle gängigen nichtlinearen Modelle, wie z.B. beim Logit-, Probit-, Tobit-, oder beim Poissonzählmodellen auf. So besitzt das in Abschnitt 3 angesprochene Probitmodell folgenden Erwartungswert gegeben den fixen Effekt und die Kontrollvariablen:

$$E(Y_t | D_t = d_t, X_t = x_t, C = c) = \Phi\left(\frac{d_t \delta_0 + x_t \beta_0 + c}{\sigma_{E_t}}\right). \quad (31)$$

---

<sup>21</sup> Für eine vertiefte Diskussion der notwendigen Annahmen aus der Perspektive der Kausalanalyse (statt aus der Perspektive eines gegebenen Panelmodells) und sich daraus ergebenden Panelmodelle siehe die wegweisenden Arbeiten von Heckman und Robb (1986) und Heckman und Hotz (1989), sowie die nichtparametrischen Erweiterungen (*Matching*), die beispielsweise in Eichler und Lechner (2001) angewendet und in Lechner (2001 d) ausführlich dargestellt werden.

Es ist offensichtlich, dass Differenzen über die Zeit von dem in Gleichung (31) angegebenen Erwartungswert immer noch von  $c$  abhängen. Kann man nicht mit Hilfe vieler Zeitperioden  $c$  konsistent schätzen ( $T \rightarrow \infty$ ), so gibt es bis jetzt keine anderen Möglichkeiten, konsistente Schätzungen der Parameter  $\delta_0$  und  $\beta_0$  zu erhalten.<sup>22</sup> Es gibt jedoch Beispiele für die einige Modelle, die es aufgrund ihrer spezifischen Annahmen über die funktionale Form des bedingten Erwartungswertes, gegeben die Regressoren und den Individualeffekt, erlauben, Momentenbedingungen abzuleiten, die nicht vom Individualeffekt abhängen und zur Identifikation von  $\delta_0$  und  $\beta_0$  ausreichen. Chamberlain (1980, 1984) zeigte dies beispielsweise für das binäre Logitmodell, das durch folgenden bedingten Erwartungswert charakterisiert ist:<sup>23</sup>

$$E(Y_t | D_t = d_t, X_t = x_t, C = c) = [1 + e^{-d_t \delta_0 - x_t \beta_0 - c}]^{-1}. \quad (32)$$

Konditioniert man jedoch noch zusätzlich auf das Ereignis, dass zwischen zwei Perioden ein Statuswechsel in der binären abhängigen Variablen stattgefunden hat ( $Y_t + Y_\tau = 1$ ), so erhält man folgende Momentenbedingung, die von unbeobachtbaren Variablen unabhängig ist und zur konsistenten Schätzung der skalierten Parameter auf der Unterstichprobe der Personen, deren binäre abhängige Variablen in den beiden betrachteten Perioden verschieden ist, verwendet werden kann:

$$E[Y_t - (1 + e^{(d_\tau - d_t)\delta_0 + (x_\tau - x_t)\beta_0})^{-1} | D_t = d_t, D_\tau = d_\tau, X_t = x_t, X_\tau = x_\tau, Y_t + Y_\tau = 1] = 0. \quad (33)$$

<sup>22</sup> Heckman (1981) ist die meistzitierte Quelle bezüglich der Verzerrung eines Probitmodells mit fixen Effekten. Arellano (2001) enthält weitergehende analytische Betrachtungen über die 'Natur' der Verzerrung. Laisney und Lechner (2001) schlagen eine 'fast' konsistente Schätzung der Modellparameter vor. Chamberlain (1980, 1984) schlägt eine sparsame Parametrisierung der Abhängigkeit von  $C$  von den übrigen Modellparametern vor. Ist diese Parametrisierung korrekt, kann das Modell im Prinzip wie ein nichtlineares Modell mit zufälligen Effekten geschätzt werden. Manski (1987) zeigt die Identifikation eines semiparametrischen Modells der binären Wahl. Für dieses Modell existiert jedoch zum einen kein Schätzverfahren, das mit Konvergenzrate  $\sqrt{N}$  gegen eine Normalverteilung konvergiert (Horowitz, 1992, Charlier, 1997, lösen dieses Problem 'fast'), und zum anderen verlangt es eine sehr ungewöhnliche zusätzliche Identifikationsbedingung.

<sup>23</sup> Zur Vereinfachung der Notation sei angenommen, die Varianzen der Fehlerterme seien  $\pi^2/3$ . Essentiell für das Folgende ist jedoch nur die Annahme, dass sie für mindestens zwei Perioden gleich sind.

Für andere Transformationen, die in verschiedenen semiparametrischen Poisson- und Tobitmodellen möglich sind, sei der Leser auf die schon erwähnte Übersichtsarbeit von Arellano und Honoré (2001) hingewiesen.<sup>24</sup>

Es wurde in diesem Beitrag schon mehrfach darauf hingewiesen, dass unter den Annahmen eines zeitkonstanten unbeobachtbaren Selektivitätseffekts folgende Differenz den kausalen Effekt wiedergibt:

$$\gamma_0(x_t) = E_{C|X_t=x_t} [E(Y_t | X_t = x_t, C = c, D_t = 1) - E(Y_t | X_t = x_t, C = c, D_t = 0)]. \quad (34)$$

Für den linearen Fall ist es unproblematisch, dass die einzelnen Komponenten der Differenz nicht identifiziert sind, denn die Differenz ist trotzdem identifiziert, da der jeweils nicht identifizierte Teil ( $c$ ) in beiden Komponenten additiv erscheint und sich aufhebt. Dies ist jedoch in nichtlinearen Modellen, in denen  $c$  nicht in additiv separabler Weise eingeht, im allgemeinen nicht der Fall. In dem hier betrachteten Logitmodell ist die Identifikation und konsistente Schätzung von  $\delta_0$  und  $\beta_0$  **nicht** ausreichend um ohne weitere Annahmen eine konsistente Schätzung des kausalen Parameters konstruieren zu können. Also haben die absoluten Koeffizientenwerte keine direkte kausale Interpretation! Jedoch entspricht das Vorzeichen von  $\delta_0$  in beiden Fällen dem Vorzeichen des von dem Modell implizierten kausalen Effekts. Aus den folgenden Umformungen erkennt man, dass sich immerhin aus der Schätzung  $\hat{\delta}_N$  eine Information über das Vorzeichen des Effekts gewinnen lässt:

$$\begin{aligned} \gamma_0(x_t) &= E_{C|X_t=x_t} [E(Y_t | X_t = x_t, C = c, D_t = 1) - E(Y_t | X_t = x_t, C = c, D_t = 0)] \\ &= E_{C|X_t=x_t} \{ [1 + e^{-d_t \delta_0 - x_t \beta_0 - c}]^{-1} - [1 + e^{-x_t \beta_0 - c}]^{-1} \} = E_{C|X_t=x_t} \frac{1 + e^{-x_t \beta_0 - c} - 1 - e^{-\delta_0 - x_t \beta_0 - c}}{(1 + e^{-\delta_0 - x_t \beta_0 - c})(1 + e^{-x_t \beta_0 - c})} \end{aligned}$$

<sup>24</sup> Das dieser Arbeit zugrundeliegende Diskussionspapier enthält auch die expliziten Ausdrücke für das

$$= (1 - e^{-\delta_0}) \underbrace{E_{C|X_t=x_t} \frac{e^{-x_t \beta_0 - c}}{(1 + e^{-\delta_0 - x_t \beta_0 - c})(1 + e^{-x_t \beta_0 - c})}}_{>0}. \quad (35)$$

Ausserdem ist die  $t$ -Statistik für das Schätzverfahren  $\hat{\delta}_N$  ein gültiger Test der Nullhypothese, dass der kausale Effekt null sei.

## 5 Schlussfolgerung

Paneldaten gewinnen als Datenmaterial für empirische Studien im Bereich der Wirtschaftswissenschaften, sowie auch in vielen anderen empirisch ausgerichteten Forschungsfeldern, eine immer grössere Bedeutung. Das Papier gab einen Überblick über einige in der ökonometrischen Paneldatenanalyse populäre lineare und nichtlineare Modelle und diskutiert die Annahmen, unter denen die geschätzten Modellparameter kausal interpretiert werden können. Es wurde dabei offensichtlich, dass die dafür notwendigen (oft impliziten) Annahmen in vielen Anwendungen als wenig plausibel erscheinen dürften. Daraus ist wahrscheinlich auch der heute in der sich explizit mit kausaler Analyse befassenden Literatur zu erkennende Trend erklärbar. Dieser Trend scheint weg von parametrischen *Panelmodellen* zu führen. Stattdessen wird die direkte (und oft nichtparametrische) Verwendung der Panelinformation postuliert, um dadurch in geeigneter Weise Selektion auf Basis beobachtbarer und unbeobachtbarer Faktoren in den Griff zu bekommen. Ziel ist es somit glaubwürdigere empirische Studien zu schaffen. Ohne Paneldaten wäre dieses Ziel noch in weiter Ferne.

## Literaturverzeichnis

Anselin, L., und A. K. Bera (1997): "Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics", in A. Ullah und D. E. A. Giles (Hrsg.), *Handbook of Applied Economic Statistics*, New York: Marcel Dekker.



- Arellano, M. (2001): "Discrete Choice with Panel Data", *mimeo*.
- Arellano, M. (2002), *Panel Econometrics*, im Erscheinen.
- Arellano, M., und B. Honoré (2001): "Panel Data Models: Some Recent Developments", wird erscheinen in J. J. Heckman und E. Leamer (Hrsg.), *Handbook of Econometrics*, Vol. V, Amsterdam: North-Holland.
- Avery, R., L. Hansen, und V. Hotz (1983): "Multiperiod Probit Models and Orthogonality Condition Estimation", *International Economic Review*, 24, 21-35.
- Baltagi, B. H., und C. Kao: "Panel Time Series Models: A Survey," *Advances in Econometrics*, Band 15, im Erscheinen.
- Baltagi, B. H. (1995), *Econometric Analysis of Panel Data*, Chichester: Wiley.
- Bertschek, I., und M. Lechner (1998): "Convenient Estimators for the Panel Probit Model", *Journal of Econometrics*, 87, 329-371.
- Chamberlain, G. (1980): "Analysis of Covariance with Qualitative Data", *Review of Economic Studies*, 47, 225-238.
- Chamberlain, G. (1984): "Panel Data", in Z. Griliches und M. D. Intriligator (Hrsg.), *Handbook of Econometrics*, Vol. II, Kap. 22, Amsterdam: North-Holland.
- Charlier, E. (1997), *Limited Dependent Variable Models for Panel Data*, PhD-Thesis, Tilburg University.
- Eichler, M. und M. Lechner (2001): "An Evaluation of Public Employment Programmes in the East German State of Sachsen-Anhalt." Wird erscheinen in *Labour Economics – An International Journal*.
- Greene, W. (2001): "Fixed and Random Effects in Nonlinear Models", *mimeo*.
- Guilkey, D. K., und J. L. Murphy (1993): "Estimation and Testing in the Random Effects Probit Model", *Journal of Econometrics*, 59, 301-317.
- Heckman, J. J., R. J. LaLonde und J. A. Smith (1999): "The Economics and Econometrics of Active Labor Market Programs", in O. Ashenfelter and D. Card (eds.): *Handbook of Labor Economics*, Vol. III A, 1865-2097, Amsterdam: North-Holland.
- Heckman, J.J. (1981): "Statistical Models for Discrete Panel Data", in C.F. Manski und D. McFadden (Hrsg.), *Structural Analysis of Discrete Data*, Cambridge: MIT Press.
- Heckman, J. J., und R. Robb (1985): "Alternative Methods of Evaluating the Impact of Interventions", in J. J. Heckman und B. Singer (Hrsg.), *Longitudinal Analysis of Labour Market Data*. New York: Cambridge University Press, 156-245.

- Heckman, J. J., und V. J. Hotz (1989): "Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training." *Journal of the American Statistical Association*. 84: 862-880.
- Horowitz, J. L. (1992): "A Smoothed Maximum Score Estimator for the Binary Response Model", *Econometrica*, 60, 505-531.
- Hsiao, C. (1986), *The Analysis of Panel Data*, Cambridge: Cambridge University Press.
- Hujer, R., U. Rendtl und G. Wagner (1997), *Wirtschafts- und Sozialwissenschaftliche Panelstudien*, Vandenhoeck & Ruprecht.
- Imbens, G. W. (2000): "The Role of the Propensity Score in Estimating Dose-Response Functions", *Biometrika*, 87, 706-710.
- Inkmann, J. (2000): "Misspecified heteroscedasticity in the panel probit model: A small sample comparison of GMM and SML estimators", *Journal of Econometrics*, 97, 227-259.
- Krishnakumar, J. (1999): "Panel Data Models", *The Current State of Economic Science*, Vol. 1, 325-344.
- Laisney, F., und M. Lechner (2001): "Almost consistent estimation of the fixed effects panel probit model", *mimeo*.
- Lechner, M. (2001a): "Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption", in M. Lechner und F. Pfeiffer (Hrsg.), *Econometric Evaluation of Labour Market Policies*, Heidelberg: Physica, 2001, 43-58.
- Lechner, M. (2001b): "Programme Heterogeneity and Propensity Score Matching: An Application to the Evaluation of Active Labour Market Policies", wird erscheinen in *The Review of Economics and Statistics*.
- Lechner, M. (2001c): "Some practical issues in the evaluation of heterogeneous labour market programmes by matching methods", wird erscheinen in *Journal of the Royal Statistical Society, Series A, Statistics in Society*.
- Lechner, M. (2001d), *Lecture notes on econometric evaluation methods*. Erhältlich unter [www.siaaw.unisg.ch/lechner](http://www.siaaw.unisg.ch/lechner).
- Lechner, M. and J. Breitung (1996): "Some GMM Estimation Methods and Specification Tests for Nonlinear Models", in L. Matyas und P. Sevestre (Hrsg.), *The Econometrics of Panel Data*, 2. Auflage, Kluwer, 583-612.
- Lee, M-J. (2002), *A Course in Panel Econometrics*. Im Erscheinen.
- Manski, C.F. (1987): "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data", *Econometrica*, 55, 357-362.

Mátyás, L., und P. Sevestre (Hrsg, 1996), *The Econometrics of Panel Data*, 2. Auflage, Dordrecht: Kluwer.

Philips, P. C. B., und H. R. Moon (1999): "Linear Regression Limit Theory for Nonstationary Panel Data", *Econometrica*, 67, 1057-1111.

Rubin, D.B. (1974): "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies", *Journal of Educational Psychology*, 66, 688-701.

van den Berg, G. J. (2001): "Duration Models: Specification, Identification, and Multiple Durations". wird erscheinen in J. J. Heckman und E. Leamer (Hrsg.), *Handbook of Econometrics*, Vol. V, Amsterdam: North-Holland.