Kapitel 10

Endogene Regressoren

> *"Man is impelled to invent theories to account for what happens in the world. Unfortunately, he is not quite intelligent enough, in most cases, to find correct explanations. So that when he acts on his theories, he behaves very often like a lunatic."*
> (Aldous Huxley, *Texts and Pretexts*, 1932, p. 270)

Angeblich leben regelmäßige Kirchenbesucher im Durchschnitt länger als ihre weniger frommen Zeitgenossen. Aber leben sie länger, *weil* sie regelmäßig zur Kirche gegangen sind? Ebenso sagt man den Bewohnern von Kreta eine höhere Lebenserwartung nach, und vielfach wird dies mit den dortigen Ernährungsgewohnheiten begründet. Aber können wir tatsächlich davon ausgehen, dass diese Kreta-Diät die *Ursache* für die höhere Lebenserwartung ist?

10.1 Kausalität

> *"The law of causality, I believe, [...] is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm."*
> (Bertrand Russel)
> (`http://www.readbookonline.net/readOnLine/22891/`)

Die Frage nach der 'Kausalität' beschäftigt die Philosophen mindestens seit Aristoteles, und in manchem fühlt man sich in den Diskussionen um das 'Wesen' der Kausalität an die Schwierigkeiten erinnert, die der Heilige Augustinus mit der *'Zeit'* hatte; in seinen Confessiones schreibt er: *"Was also ist die Zeit? Wenn mich niemand danach fragt, weiß ich es. Wenn ich es einem erklären will, der danach fragt, weiß ich es nicht"* (Augustinus 1980, Liber XI, Caput XIV).

In der Ökonomik steht die Kausalität als fundamentales Konzept ganz am Anfang, bekanntlich trägt Adam Smiths epochales Werk den Titel *"An Inquiry into the Nature and Causes of the Wealth of Nations"* (1776).

Im ersten Moment scheint es klar und offensichtlich, was eine 'Ursache – Wirkungs' Beziehung ist, mein gestriger Alkoholkonsum ist die Ursache für mein heutiges Kopfweh. Aber war es wirklich der Alkohol? Könnte nicht das übermäßige Rauchen oder der kurze Schlaf die Ursache sein, oder eine sich ankündigende Grippe?

Während eine Korrelation nur eine Beschreibung des Zusammenhangs zweier Variablen liefert (Assoziation), sucht die Kausalität nach Ursache und Wirkung, versucht eine Erklärung zu geben. Wir beobachten zwar, dass Lottogewinner häufiger Fernurlaube machen (Korrelation), aber einen Fernurlaub anzutreten, um im Lotto zu gewinnen, könnte schief gehen (Kausalität).

Die heute in den Sozialwissenschaften verbreitete Vorstellung von Kausalität wurde wesentlich von David Hume (1711 – 1776) geprägt. Er stellt fest, dass wir Kausalität nicht unmittelbar sinnlich wahrnehmen können, sondern lediglich die Abfolge von Ereignissen beobachten. Er schreibt

> "Wenn aber viele gleichförmige Beispiele auftreten und demselben Gegenstand immer dasselbe Ereignis folgt, dann beginnen wir den Begriff von Ursache und Verknüpfung zu bilden. Wir empfinden nun ein neues Gefühl [...]; und dieses Gefühl ist das Urbild jener Vorstellung [von notwendiger Verknüpfung], das wir suchen." (Hume, 1993, 95) [zitiert nach `https://de.wikipedia.org/wiki/Kausalit%C3%A4t`].

Wir können zwar beobachten, dass sich ein Stein im Sonnenschein erwärmt, und dass diese Phänomene sehr häufig parallel auftreten, aber unsere unmittelbare Sinneswahrnehmung erlaubt uns noch keine Rückschlüsse auf dahinter liegende Ursachen-Wirkungs Beziehungen.

Die Verknüpfung von Ursache und Wirkung ist demnach lediglich eine menschliche Gewohnheit und keine Eigenschaft der beobachteten Objekte. Nach Hume können wir lediglich räumlich benachbarte Ereignisse in zeitlicher Abfolge beobachten. Die Erfahrung der Regularität (*'constant conjunction'*) als Ursachen und Wirkung *erleben* wir zwar häufig als kausal, aber wie bei Induktionsschlüssen können aus beobachteten Regelmäßigkeiten nie absolut sichere Schlussfolgerungen über zukünftige Ereignisse ziehen, solche Schlussfolgerungen sind bestenfalls (sehr) wahrscheinlich.

Der Versuch die Begriffe Kausalität und Ursache präziser in Form notwendiger und hinreichender Bedingungen zu fassen führt zur so genannten INUS Konzeption des australischen Philosophen John Mackie, siehe Exkurs (Seite 3).

Beispiel: *Verursachen* höhere Werbeausgaben x höhere Umsätze y?

Dazu überlegen wir, welche mögliche Ursachen für eine beobachtete Korrelation zwischen Werbeausgaben x Umsätze y denkbar sind:

- x ist die Ursache für y: höhere Werbeausgaben führen zu höheren Umsätzen.
- y ist die Ursache für x: (*reverse causality*) höhere Umsätze ermöglichen die Finanzierung zusätzlicher Werbeausgaben.
- z ist eine gemeinsame Ursache für x und y: (*confounding variable*, Scheinkorrelation), z.B. eine gute Konjunktur führt zu steigenden Umsätzen *und* zu steigenden Werbeausgaben.

Exkurs: Die INUS Konzeption

Der Versuch, *Kausalität* in den Begriffen notwendiger und hinreichender Bedingungen zu erfassen, führt zur neueren INUS Konzeption von Kausalität (*"insufficient, but necessary part of an unnecessary but sufficient condition"*).
Angenommen ein spezielles Nahrungsmittel führe bei Menschen mit einer speziellen genetischen Prädisposition immer zu Krebs. Nach der INUS Konzeption gilt

- die Einnahme des Nahrungsmittels ist kein hinreichender Teil der Bedingung "Nahrungsmittel plus genetische Disposition", weil die Einnahme alleine nicht zwangsläufig zu Krebs führt;
- die Einnahme des Nahrungsmittels ist aber ein notwendiger Teil der Bedingung, weil sonst die Bedingung "Nahrungsmittel plus genetische Disposition" nicht erfüllt wäre;
- "Nahrungsmittel plus genetische Disposition" ist eine hinreichende Bedingung für die Krebserkrankung, weil sie zwangsläufig zur Krebserkrankung führt (bzw. den Erwartungswert erhöht);
- aber "Nahrungsmittel plus genetische Disposition" ist keine notwendige Bedingung für die Krebserkrankung, weil die Krebserkrankung auch durch andere Bedingungen verursacht werden kann;

Das Nahrungsmittel kann genau dann als Ursache von Krebs betrachtet werden, wenn die Einnahme des Nahrungsmittels allein kein hinreichender, aber dennoch notwendiger Teil einer Bedingung ist, die selbst hinreichend, aber nicht notwendig für die Krebserkrankung ist.

¶

- Die Korrelation zwischen x und y tritt zufällig auf: (dies – und nur dies – sollte durch statistische Tests erkennbar sein)

□

Auf Hume gehen auch die zwei zentralen Ideen zurück, die insbesondere für empirische Analysen bis heute von zentraler Bedeutung sind, er schreibt:

> *"We may define a cause to be an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second. Or, in other words, where, if the first object had not been, the second never had existed."*
> (An Enquiry concerning Human Understanding, Section VII)

Diese zwei Ideen sind

1. die zeitliche Anordnung: die Wirkung kann nicht *vor* der Ursache eintreten; wie die Zeit ist Kausalität asymmetrisch. Wir beobachten, dass Reiche häufig Luxusurlaube machen; aber es wäre keine gute Idee, einen Luxusurlaub zu machen, um reich zu werden.

In der Ökonometrie führte die Ausnützung der zeitlichen Asymmetrie von Ursache-Wirkungsbeziehungen in Zeitreihen zum Konzept der *'Granger Kausalität'*, gemäß dem lateinischen *'post hoc, ergo propter hoc'* ('danach, also deswegen'), die Clive Granger den Alfred-Nobel-Gedächtnispreis für Wirtschaftswissenschaften einbrachte.

Beispiel: Angenommen wir interessieren uns dafür, ob höhere Werbeausgaben W höhere Umsätze S (für *'sales'*) generieren, oder umgekehrt, höhere Umsätze die Finanzierung höherer Werbeausgaben ermöglichen, könnte man die folgenden Gleichungen schätzen

$$
\begin{aligned}
W_t &= \alpha + \sum_{i=1}^{I} \beta_i W_{t-i} + \sum_{j=1}^{J} \gamma_j S_{t-j} + \epsilon_t \\
S_t &= \delta + \sum_{k=1}^{K} \zeta_k S_{t-k} + \sum_{l=1}^{L} \lambda_l W_{t-l} + \nu_t
\end{aligned}
$$

wobei W und S stationäre Zeitreihen sind und ϵ und ν nur weißes Rauschen enthalten sollen. Dies Gleichungen können mittels OLS geschätzt werden.

Wenn in der ersten Gleichung die Nullhypothese, dass alle γ_j simultan gleich Null sind, verworfen werden kann sagt man, höhere Umsätze S *Granger-verursachen* höhere Werbeausgaben W.

Wenn umgekehrt in der zweiten Gleichung die Nullhypothese, dass alle λ_l simultan gleich Null sind, verworfen werden kann sagt man, höhere Werbeausgaben W Granger-verursachen höhere Umsätze S.

Wenn sowohl alle γ_j als auch alle λ_l gemeinsam signifikant von Null verschieden sind spricht man von 'feed-backs'.

Obwohl auf den ersten Blick sehr einleuchtend sollte man mit diesem Konzept vorsichtig umgehen. (Leamer, 1985, 259, 283) warnt z.B.: "this concept should be called 'precedence' [...] We can think all contexts in which precedence is suggestive of causation and also contexts in which it is not [...]. It is altogether clear that precedence is not sufficient for causality. Weather forecasts regularly precede the weather, but few of us take this as evidence that the forecasts 'cause' the weather".

2. die Vorstellung einer *kontrafaktischen Situation*: wäre das verursachende Ereignis *nicht* eingetreten, wäre die Folge ausgeblieben.

   Die zentrale Idee der kontrafaktischen Situationen (Kontrafaktizität) wurde u.a. von Lewis (1973) weiterentwickelt. Diese Konzeption ermöglicht eine Unterscheidung von Kausalität von bloßer zeitlicher Abfolge.

   Vergleiche Opportunitätskosten.

   Allerdings existiert die Welt (wahrscheinlich) nur in einer Ausgabe, deshalb können wir nur für jeden einzelnen Fall beobachten, ob ein Ereignis A eingetreten ist oder nicht, und ob darauf B gefolgt ist oder nicht, aber wir können nicht die Situation und die kontrafaktische Situation parallel beobachten. Dies ist die alte *"was wäre wenn"* Frage; wir können z.B. nicht wissen, wie die Geschichte des 20. Jahrhunderts ausgesehen hätte, *wenn* die Kunstakademie Wien den jungen Hitler aufgenommen hätte.

In der empirischen Forschung wurde die kontrafaktische Theorie der Kausalität v.a. im Zusammenhang mit einer *interventionistischen Theorie der Kausalität* (Gaskin 1955, von Wright 1971) sehr einflussreich. Dabei wird der Begriff der 'Verursachung' mit der Manipulation von Dingen verknüpft. Die dahinter liegende Vorstellung ist, dass die Manipulation einer Ursache eine Wirkung hervorrufen wird, oder in anderen Worten, wenn zwei Ereignisse A und B kausal verknüpft sind, dann kann durch Manipulation von A das Ereignis B beeinflusst werden. Damit wird der Kausalitätsbegriff auf aktive Handlungen zurückgeführt.

Der zentrale Begriff dabei ist der einer *Intervention*, ein gewissermaßen 'chirurgischer' (isolierter) Eingriff auf Ereignis A der derartig gestaltet ist, dass eine Veränderung in B ausschließlich auf die Manipulation von A zurückgeführt werden kann (für eine ausführliche Diskussion siehe Woodward (2016)).

Damit wird auch verständlich, weshalb diese Konzeption v.a. in den Sozialwissenschaften und in der Medizin solche Bedeutung erlangte, es geht um aktive Eingriffe in die Natur oder Gesellschaft um Zustände gezielt zu verändern.

Naturwissenschaftliche Experimente beruhen auf solchen systematischen Interventionen, mit deren Hilfe wir Wissen über kausale Zusammenhänge zu erlangen suchen.

10.1.1 Experimente

> *"... no one believes an hypothesis except its originator but everyone believes an experiment except the experimenter."*
> (W.I.B. Beveridge, 1950, p65)

Als einer der Begründer naturwissenschaftlicher Experimente gilt Galileo Galilei (1564 – 1642).

Mit Hilfe seiner Experimente auf der schiefen Ebene, siehe Abbildung 10.1, konnte er eine mathematische Beschreibung der Fallgesetze finden.

Ein ist im allgemeinen durch die zwei folgenden wesentlichen Elemente gekennzeichnet:

1. Kontrolle
2. Intervention

Zeitmessung mit Wasseruhren und eigenem Puls.

Mathematische Beschreibung der Resultate.

Abbildung 10.1: Galileo Galilei: Versuche auf der schiefen Ebene, historische Darstellung und Nachbau.

> "Die Philosophie steht in diesem großen Buch geschrieben, dem Universum, das unserem Blick ständig offen liegt. Aber das Buch ist nicht zu verstehen, wenn man nicht zuvor die Sprache erlernt und sich mit den Buchstaben vertraut gemacht hat, in denen es geschrieben ist. Es ist in der Sprache der Mathematik geschrieben, und deren Buchstaben sind Kreise, Dreiecke und andere geometrische Figuren, ohne die es dem Menschen unmöglich ist, ein einziges Wort davon zu verstehen; ohne diese irrt man in einem dunklen Labyrinth herum." (Galileo Galilei: Il Saggiatore, zitiert aus `https://de.wikipedia.org/wiki/Galileo_Galilei`)

Jede einzelne Durchführung gleicht einem induktiven Schluss, aber wenn das Experiment von vielen Personen wiederholt werden kann steigt, nach Popper, der Grad der Bewährung.

Außerhalb der Naturwissenschaften ist eine perfekte Kontrolle kaum möglich.
Zwei Alternativen:

1. 'Nachbau' der Realität in Form eines mathematischen Modells und hypothetische Interventionen unter ceteris paribus Annahme (komparative Statik, Multiplikatoranalyse): Cowles Commission Approach

2. Experimente mit Randomisierung: *Randomized Controlled Trials* (RCT): wenn nicht alle relevanten Faktoren 'kontrolliert' werden können (bzw. nicht a priori bekannt ist, welche Faktoren relevant sind), sorgt Randomisierung dafür, dass keine *systematischen* Unterschiede zwischen den Gruppen (Interventions- und Kontrollgruppe) auftreten.

   In der Sprache der Ökonometrie: bei wiederholter Durchführung des Experiments sollte sich eine Korrelation zwischen Störtermen und Regressoren 'wegmitteln' (für eine einzelne Durchführung kann nicht damit gerechnet werden, dass die Stichprobenkorrelation gleich Null ist!).

Fisher's Feldversuche in Rothamsted: die einzelnen Plots unterscheiden sich nach Bodenqualität, Feuchtigkeit, Sonneneinstrahlung etc. Wenn keine perfekte Kontrolle möglich ist: Randomisierung!

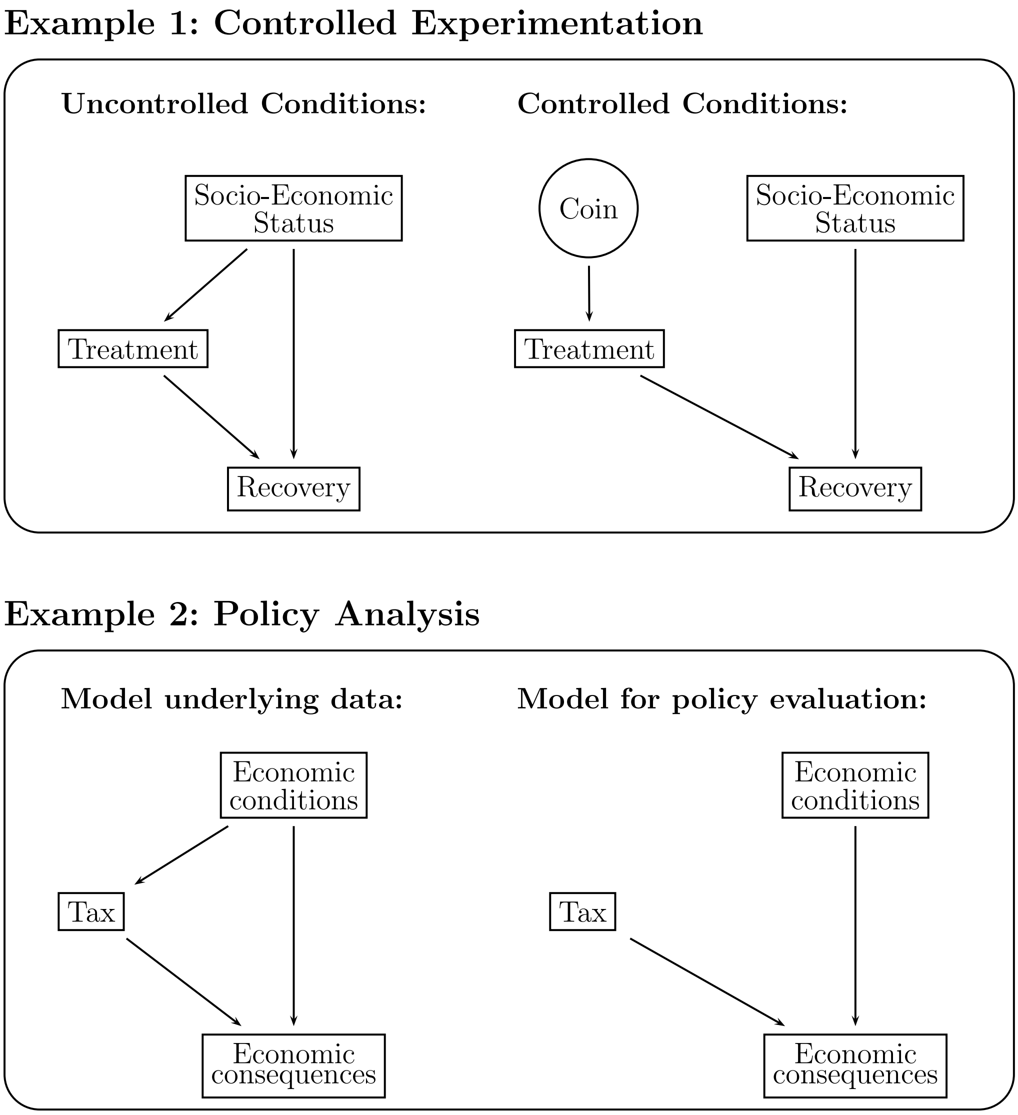


Abbildung 10.2: Kontrollierte und nicht kontrollierte Experimente (entnommen aus: Pearl (2000, 347f).

Abbildung 10.2 veranschaulicht das Problem.

Kritik: Deaton and Cartwright (2016)

10.2 Validität

Im Zusammenhang mit Kausalanalysen bezieht sich der Begriff der Validität (*'validity'*) auf die Gültigkeit und Übertragbarkeit von Schlußfolgerungen. Da induktive Schlüsse nie bewiesen werden können, kann auch Validität nicht im strengen Sinne bewiesen werden, sie muß jeweils argumentativ begründet werden. Deshalb spricht man besser vom *Validitätsgrad* einer Schlußfolgerung oder Methode.

Generell unterscheidet man zwischen interner und externer Validität. Eine statistische Analyse ist *intern valid*, wenn die Schlußfolgerungen aus der Stichprobe für die untersuchte Grundgesamtheit gültig sind, und wenn alternative Erklärungsmöglichkeiten weitestgehend ausgeschlossen werden können. Eine Analyse ist *extern valid*, wenn die Schlußfolgerungen auf andere Grundgesamtheiten oder Umstände übertragen lassen.

Wird zum Beispiel eine Absolventenbefragung an der Universität Innsbruck durchgeführt, so geht es bei der internen Validität darum, wie gut mit den zur Verfügung stehenden Methoden aus den beantworteten Fragebögen auf die Grundgesamtheit aller Studierenden der Universität Innsbruck geschlossen werden kann. Bei der externen Validität geht es im nächsten Schritt darum, inwieweit die Ergebnisse für die Innsbrucker Studierenden auf Studierende anderer Universitäten übertragbar sind.

10.2.1 Interne Validität

Interne Validität hat zumindest zwei Komponenten, erstens, inwieweit der Schätzer für den kausalen Effekt unverzerrt und konsistent ist, und zweitens, inwieweit Hypothesentests zuverlässige Ergebnisse liefern, das heißt, inwieweit die Standardfehler richtig geschätzt wurden und die Verteilungsannahmen zutreffend sind.

Wie bereits früher gezeigt führt eine Korrelation der Störterme mit den unabhängigen x Variable(n)[1], zu verzerrten und nicht konsistenten OLS Schätzern für die Parameter. Solche Fälle, in denen die interne Validität verletzt ist, treten z.B. auf bei

- fehlende relevante x Variablen (*'omitted variable bias'*);
  mögliche Maßnahmen: a) Instrumentvariablen, b) Proxy Variablen, c) Randomisierung
- Fehlspezifikation in Bezug auf die Funktionsform;
  mögliche Maßnahmen: Transformation der Variablen, nicht-lineare Schätzmethoden
- Messfehler in den x Variablen (*'errors-in-variables bias'*);
  mögliche Maßnahmen: a) Instrumentvariablen, b) Modellierung des Meßfehlers

[1] Allgemeiner, wenn der auf die x bedingte Erwartungswert der Störterme ungleich Null ist, $E(u_i|x_i) \neq 0$

- Stichprobenfehler (*'sample selection bias'*);
  mögliche Maßnahmen: a) Heckit, b) Matching
- Simultane Kausalität (*'simultaneous causality bias'*);
  mögliche Maßnahmen: a) Instrumentvariablen, b) Experimente

Verzerrte und nicht konsistente Standardfehler, die die interne Validität von Hypothesentests gefährden, folgen z.B. aus Heteroskedastizität und Autokorrelation.

10.2.2 Externe Validität

Die externe Validität ist z.B. gefährdet wenn

- Unterschiede in der Grundgesamtheit, oder
- Unterschiede in den Rahmenbedingungen bestehen.

Probleme mit der externen Validität können z.B. aufgrund theoretischer Überlegungen vermutet werden, oder aufgrund eines Vergleichs von Studien für verschiedene relevante Grundgesamtheiten erkannt werden.

10.2.3 Mögliche Probleme bei der praktischen Durchführung von Experimenten

Wir bereits früher ausgeführt kann zwischen einer internen und externen Validität unterschieden werden, wobei es bei der internen Validität vor allem darum geht, wie gut die Stichprobenregressionsfunktion die Grundgesamtheit abbildet (z.B. Unverzerrtheit und Konsistenz), und bei der externen Validität darum, inwieweit die Ergebnisse auf andere Grundgesamtheiten übertragbar oder unter anderen Rahmenbedingungen gültig sind.

Mögliche Probleme für die interne Validität

- **Mangel- oder fehlerhafte Randomisierung:** Wenn zum Beispiel eine Telefonumfrage untertags im Festnetz durchgeführt wird sind vermutlich nicht Berufstätige überrepräsentiert.
- **Mangelhafte Mitwirkung der Versuchspersonen** (*failure to follow treatment protocol*, *partial compliance*): Versuchspersonen können nicht gezwungen werden z.B. alle Anordnungen exakt zu erfüllen.
- **Natürlicher Abgang** (*attrition*): nach Randomisierung können Personen ausfallen, insbesondere bei länger dauernden Experimenten (z.B. Arbeitsmarktmaßnahmen), und dieser Abgang erfolgt meist nicht zufällig.

- **Hawthorne und John Henry Effekte:** Personen, die sich beobachtet fühlen, ändern manchmal ihr Verhalten. Die Bezeichnung leitet sich von den Hawthorne Werken (General Electric Company, in der Nähe von Chicago) her, un denen von 1924-1932 Produktivitätsstudien durchgeführt wurden. Da die Teilnehmer wußten, dass sie in der *treatment* Gruppe waren beobachtet wurden, lag die Vermutung nahe, dass sie dies zu vermehrtet Arbeit anspornen würde.[2] Während sich der Hawthorne Effekt auf Verhaltensänderungen der *treatment* Gruppe bezieht, versteht man unter John Henry Effekten Verhaltensänderungen bei der Kontrollgruppe (z.B. aus Frustration nicht in den Genuß einer Behandlung zu kommen).
- **Kleine Stichproben:** Experimente sind häufig teuer, und deshalb die Stichproben oft klein.

Mögliche Probleme für die externe Validität

- **Nicht repräsentative Stichprobe:** wenn zum Beispiel Freiwillige Teilnehmer für ein Experiment ausgewählt werden werden ist fraglich, ob dies nicht eine Stichprobenverzerrung nach sich zieht.
- **Nicht repräsentative experimentelle Maßnahmen:** zum Beispiel Dauer eines Experiments kann zu kurz sein, am langfristige Verhaltensänderungen zu messen; auch Größeneffekte können eine Rolle spielen.
- **Gleichgewichtseffekte:** Experimente werden fast immer in einem kleinen Rahmen durchgeführt. Wird zum Beispiel ein Arbeitsmarktprogramm, dessen Auswirkungen vorher im kleinen Rahmen getestet wurden, flächendeckend eingeführt, kann dies Auswirkungen auf das allgemeine Gleichgewicht haben und *feed-backs* bewirken, die experimentell nicht vorhersehbar waren.

10.2.4 Beobachtungsstudien

Experimente häufig nicht durchführbar oder ethisch nicht vertretbar.

Alte Debatte: *Verursacht Rauchen Lungenkrebs?* (→ Epidemiologie)

Offensichtlich ist Rauchen keine notwendige Bedingung für die Entstehung der Krankheit, denn es erkranken auch viele Nichtraucher; Krebs kann viele Ursachen haben. Andererseits ist Rauchen auch keine hinreichende Bedingung, denn nicht alle Raucher erkranken tatsächlich; oft sind bestimmte Kofaktoren erforderlich, damit die Wirkung eintritt, z.B. eine genetische Disposition.

Tatsächlich publizierte R.A. Fisher in seinen letzten Lebensjahren mehrere Papers (u.a. im 'Nature'), in denen er bezweifelte, dass die bis dahin vorliegende empirische Evidenz den Schluss zulasse, dass eine kausaler Zusammenhang zwischen Rauchen und Lungenkrebs bestehe (für eine ausführliche Darstellung siehe Salsburg (2002, 181ff)).

---

[2] Eine nachträgliche genauere Auswertung der Hawthorne Daten ergab, dass in diesen Experimenten vermutlich *kein* Hawthorne Effekt vorlag, vgl. Stock and Watson (2006, S. 474).

Retrospektive Studien: Bereits Erkrankte werden nach Lebensgewohnheiten befragt.

Probleme: Omitted Variables (Confounding; z.B. sozio-ökonomischer Status); beruht häufig auf subjektiven Erinnerungen,

Prospektive Studien: Personen werden nach Rauchgewohnheiten befragt und dementsprechend in Gruppen eingeteilt. Nach entsprechender Zeit werden die Gruppen untersucht.

Probleme: Omitted Variables (Confounding), Selektive Stichproben (opportunity samples; externe Validität); manchmal kleine Stichproben.

10.3 Eine erste Intuition: Die Methode der Zweistufigen Kleinsten Quadrate

Stellen Sie sich vor, Sie sollten die Auswirkungen der Lernzeit (LZ) auf ein Prüfungsresultat (Pkt) untersuchen. Sie befragen dazu sechzig Studierende nach den Stunden, die sie für die Prüfungsvorbereitung verwendeten, und nach den Punkten, die sie schließlich bei der Prüfung erzielten (Pkt). Angenommen, Sie erhalten bei der Befragung die Daten, die im Scatter-Diagramm in Abbildung 10.3 dargestellt sind.

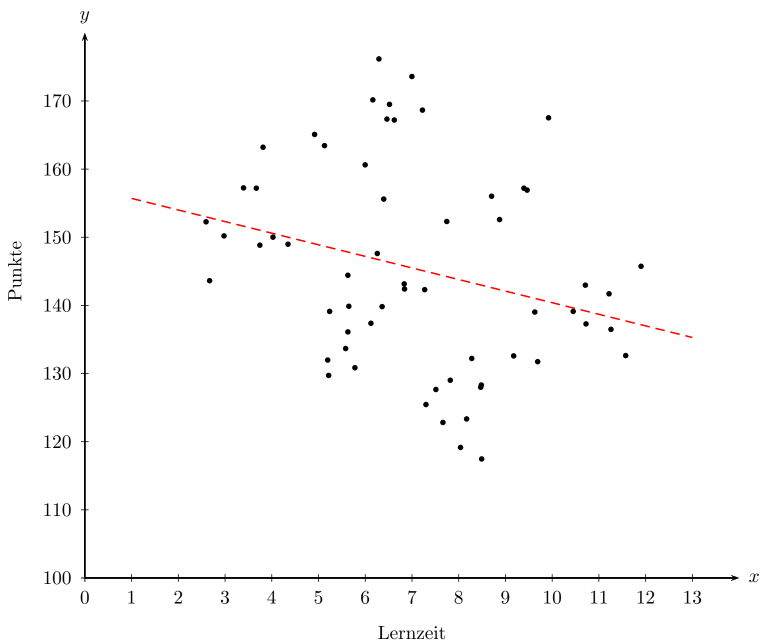


Abbildung 10.3: Zusammenhang zwischen Prüfungserfolg (Y) und Lernzeit (LZ)

Eine Regression der Punkte auf die Lernzeit gibt folgendes Ergebnis, welches strichliert in Abbildung 10.3 eingezeichnet ist.

$$\begin{array}{rcccc} \text{Pkt} & = & 157.404 & - & 1.701 \text{ LZ} \\ & & (6.004)^{***} & & (0.807)^{**} \end{array}$$

$$R^2 = 0.071, \qquad n = 60$$

Das Ergebnis scheint nahe zulegen, dass es einen negativen Zusammenhang zwischen Lernzeit und Prüfungserfolg gibt. Die Lernzeit scheint zwar nur einen relativ kleinen Anteil der Streuung zu erklären ($R^2 = 0.07$), aber der Zusammenhang ist fast auf dem 5%-Niveau signifikant von Null verschieden.

Bedeutet dies, dass Studierende, die länger lernen, bei der Prüfung mit einer schlechteren Note rechnen müssen???

Nicht unbedingt, möglicherweise hängt das Prüfungsergebnis auch von unbeobachteten Fähigkeiten wie z.B. der Intelligenz ab, und möglicherweise bereiten sich klügere Studierende im Durchschnitt weniger lange auf die Prüfung vor.

In diesem hypothetischen Beispiel haben wir einen Indikator für die Fähigkeiten (F), wenn wir diesen in der Regression berücksichtigen erhalten Sie folgendes Ergebnis

$$\begin{array}{rcccccc} \text{Pkt} & = & \underset{(7.351)}{8.278} & + & \underset{(0.421)^{***}}{5.052\,\text{LZ}} & + & \underset{(0.048)^{***}}{1.014\,\text{F}} \end{array}$$

$$R^2 = 0.895, \qquad n = 60$$

Dieses Ergebnis ist deutlich weniger überraschend, alle Koeffizienten weisen nun das erwartete Vorzeichen auf und sind hochsignifikant von Null verschieden.

Was ist passiert?

Tatsächlich wurden die Daten mit dem Computer erzeugt, und zwar für drei Gruppen von Studierenden mit unterschiedlichen F (d.h. für F $= 75, 100$ und 125), deshalb kennen wir für dieses Beispiel das 'wahre' Modell.

Konkret wurde angenommen, dass

$$\text{Pkt}_i = 10 + 5\text{LZ}_i + 1\,\text{F}_i + \varepsilon_i \qquad \text{mit } \varepsilon_i \sim \text{N}(0, 25)$$

und zusätzlich wurde ein negativer Zusammenhang zwischen Lernzeit und den Fähigkeiten F der folgenden Art unterstellt

$$\text{LZ}_i = 12 - 0.1\,\text{F}_i + \upsilon_i$$

Der datengenerierende Prozess sieht also tatsächlich anders aus als in der 'kurzen' Regression auf die Lernzeit unterstellt, er wird in diesem Fall besser durch drei unterschiedliche Regressionsgleichungen beschrieben, wie Abbildung 10.4 dargestellt.

Eine Nichtberücksichtigung der relevanten Variable Fähigkeiten würde also zu völlig falschen Schlussfolgerungen führen!

Leider ist dies kein bei den Haaren herbeigezogenes Beispiel, tatsächlich dürfte die Nichtberücksichtigung relevanter Variablen bei praktischen Arbeiten eines der schwerwiegendsten Probleme sein. Deshalb wollen wir dieses Problem nun etwas ausführlicher diskutieren.

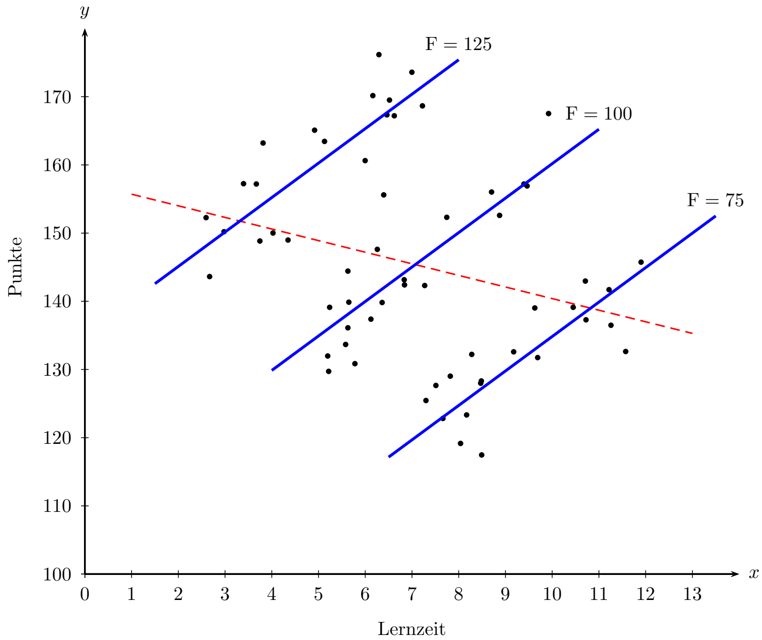


Abbildung 10.4: Zusammenhang zwischen Prüfungserfolg (Pkt) und Lernzeit (LZ) unter Berücksichtigung unterschiedlicher Fähigkeiten.

10.3.1 Nichtberücksichtigung relevanter Variablen (*Omitted Variables*)

Tatsächlich haben wir diesen Fall bereits im Rahmen der deskriptiven Regressionsanalyse ausführlich diskutiert, deshalb werden wir die Konsequenzen einer irrtümlichen Nichtberücksichtigung einer relevanten x Variable nur kurz wiederholen.

Angenommen das "wahre" Modell sei

$$\ddot{y}_i = \beta_2 \ddot{x}_{i2} + \beta_3 \ddot{x}_{i3} + \varepsilon_i$$

wobei alle Variablen in Abweichungsform dargestellt sind, d.h. $\ddot{x}_i := x_i - \bar{x}$, aber wir schätzen ein "falsches" (kurzes) Modell

$$\ddot{y}_i = \widehat{\beta}_2^* \ddot{x}_{i2} + \hat{\varepsilon}_i^*$$

Der geschätzte OLS Koeffizient des "falschen" Modells ist

$$\widehat{\beta}_2^* = \frac{\sum \ddot{x}_{i2} \ddot{y}_i}{\sum \ddot{x}_{i2}^2}$$

Um die Erwartungstreue zu überprüfen setzen wir wie üblich für $\ddot{y}_i$ das "wahre" Modell $\ddot{y}_i = \beta_2 \ddot{x}_{i2} + \beta_3 \ddot{x}_{i3} + \varepsilon_i$ ein und bilden (für deterministische x) den Erwartungswert:

$$
\begin{aligned}
\mathrm{E}(\widehat{\beta}_2^*) &= \mathrm{E}\left(\frac{\sum \ddot{x}_{i2}(\overbrace{\beta_2\ddot{x}_{i2} + \beta_3\ddot{x}_{i3} + \varepsilon_i}^{y_i})}{\sum \ddot{x}_{i2}^2}\right) \\
&= \mathrm{E}\left(\beta_2\right) + \mathrm{E}\left(\beta_3 \frac{\sum \ddot{x}_{i2}\ddot{x}_{i3}}{\sum \ddot{x}_{i2}^2}\right) + \mathrm{E}\left(\frac{\sum \ddot{x}_{i2}\varepsilon_i}{\sum \ddot{x}_{i2}^2}\right) \\
&= \beta_2 + \beta_3 \frac{\sum \ddot{x}_{i2}\ddot{x}_{i3}}{\sum \ddot{x}_{i2}^2} \\
&= \beta_2 + \beta_3 \frac{\operatorname{cov}(x_2, x_3)}{\operatorname{var}(x_2)}
\end{aligned}
$$

Man beachte, dass dies auch bei exogenen Regressoren gilt, d.h. wenn $\mathrm{E}[(x_{i2} - \bar{x}_2)\varepsilon_i] = 0$ und $\mathrm{E}(\varepsilon_i) = 0$!

In diesem Fall ist nicht die Gauss Markov Annahme A3 $\mathrm{E}(\varepsilon_i|\boldsymbol{X}) = 0$ verletzt (endogene Regressoren), sondern die Annahme A1, die Funktion ist fehlspezifiziert!

Wenn wir berücksichtigen, dass $\operatorname{cov}(x_2, x_3)/\operatorname{var}(x_2)$ der Steigungskoeffizient einer Regression der fehlenden x_3 Variable auf die vorkommende x_2 Variable, also $x_{i3} = \delta_1 + \delta_2 x_{i2} + \upsilon_i$, können wir in diesem einfachen bivariaten Fall auch schreiben

$$
\mathrm{E}(\widehat{\beta}_2^*) = \beta_2 + \beta_3\delta_2 \tag{10.1}
$$

Wir interessieren uns für den 'kausalen' Effekt β_2, die OLS Schätzung eines Modells mit fehlenden relevanten Variablen (*'omitted variables'*) liefert aber nur die gemeinsame Auswirkung $\beta_2 + \beta_3\delta_2$.

Dies ist der berühmte *'omitted variables bias*:

> Wann immer eine nicht berücksichtigte Variable mit zumindest einem im Modell vorkommenden Regressor x *und* der abhängigen Variable y korreliert ist, führt diese Nichtberücksichtigung der relevanten Variable zu systematisch verzerrten Schätzergebnissen)!
> Da dieses Problem mit zunehmender Stichprobengröße nicht kleiner wird ist eine solche OLS Schätzung auch *nicht konsistent*!

Man kann dies auch als ein Identifikationsproblem verstehen, wenn eine relevante (d.h. mit y und x_2 korrelierte) Variable fehlt ist β_2 nicht identifiziert!

Im multiplen Regressionsmodell sind die Formeln komplexer, aber auch dort führt eine einzige fehlende relevante Variable in der Regel dazu, dass *alle* Koeffizienten verzerrt und nicht konsistent geschätzt werden!

Im früheren Beispiel mit den Prüfungspunkten interessierten wir uns für den kausalen Effekt der Lernzeit $\widehat{\beta}_2 = 5.052$, durch Nichtberücksichtigung der Fähigkeiten erhielten wir einen verzerrten Koeffizienten -1.701.

Da es sich um ein hypothetisches Beispiel handelt kennen wir die wahren Werte, unter Berücksichtigung von $\mathrm{F}_i = 147.096 - 6.661\mathrm{LZ}_i + \nu_i$ können wir einfach überprüfen $\widehat{\beta}_2 + \widehat{\beta}_3 * \hat{\delta} = 5.052 + 1.014 * (-6.661) = -1.701$.

Tabelle 10.1: Gleichung (10.1) erlaubt eine Abschätzung des Vorzeichens des Bias bei der Schätzung von $\widehat{\beta}_3$, wenn x_3 in der Regression $y_i = \beta_2 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$ fälschlich nicht berücksichtigt wird.

	$\text{corr}(x_2, x_3) > 0$	$\text{corr}(x_2, x_3) < 0$
$\beta_3 > 0$	positiver Bias	negativer Bias
$\beta_3 < 0$	negativer Bias	positiver Bias

Wie Gleichung (10.1) zeigt hängt die Richtung der Verzerrung vom Vorzeichen des Koeffizienten β_3 und von der Kovarianz zwischen x_2 und x_3 ab (vgl. Tabelle 10.1).

Im Beispiel hängt die Punktezahl positiv von den unbeobachteten Fähigkeiten ab ($\beta_3 > 0$) und die Fähigkeiten sind negativ mit der Lernzeit korreliert ($\text{corr}(x_3, x_2) < 0$), deshalb wird der Effekt unterschätzt, der Bias ist negativ.

Für das Folgende ist es wichtig zu erkennen, dass diese Fehlspezifikation zu einer Korrelation mit dem Störterm führt: Wir haben irrtümlich ein kurzes Modell

$$\text{Pkt} = \widehat{\beta}_1^* + \widehat{\beta}_2^* \text{LZ} + \varepsilon^*$$

geschätzt, das wahre Modell ist aber

$$\text{Pkt} = \beta_1 + \beta_2 \text{LZ} + \underbrace{\beta_3 \, \text{F} + \varepsilon}_{\varepsilon^*}$$

Der Störterm ε^* der fehlspezifizierten Regression enthält den Einfluss aller nicht berücksichtigten Variablen, deshalb ist $\varepsilon^* = \beta_3 \text{F} + \varepsilon$, und wenn – wie in diesem Fall – F und LZ korreliert sind führt das im fehlspezifizierten Modell zu einer Korrelation zwischen dem Störterm ε^* und LZ. Deshalb liefert eine OLS Schätzung für β_2 im fehlspezifizierten Modell ein systematisch verzerrtes Ergebnis.

Eine solche Korrelation zwischen Regressoren und Störtermen wird in der Ökonometrie *Endogenität* genannt. Wir werden später sehen, dass auch simultane Kausalität oder Messfehler bei den Regressoren zu solcher Endogenität führt, und dass in all diesen Fällen die OLS Schätzer weder erwartungstreu noch konsistent sind.

Aber bereits jetzt sei betont, dass dieser ökonometrische Endogenitätsbegriff sich von dem Endogenitätsbegriff unterscheidet, der in anderen Disziplinen (z.B. Mikro- oder Makroökonomik) üblich ist.

Um eine intuitive Vorstellung von den Problemen zu bekommen erinnern wir uns, dass durch eine Regression eine zu erklärende Variable y gewissermaßen in zwei Teile zerlegt wird, in einen durch die Regressoren erklärten (systematischen) Teil, und in einen unerklärten Teil, der im Störterm abgebildet wird. Dies ist im linken Panel von Abbildung 10.5 dargestellt; wenn das Modell korrekt spezifiziert ist (!) kann der zu schätzende Koeffizient β_2 kausal interpretiert werden.

Falls aber die erklärende Variable x mit dem Störterm ε korreliert ist, wie im rechten Panel von Abbildung 10.5 dargestellt, wird der eigentlich interessierende Zusammenhang zwischen x und y durch den zusätzlichen Zusammenhang zwischen x und

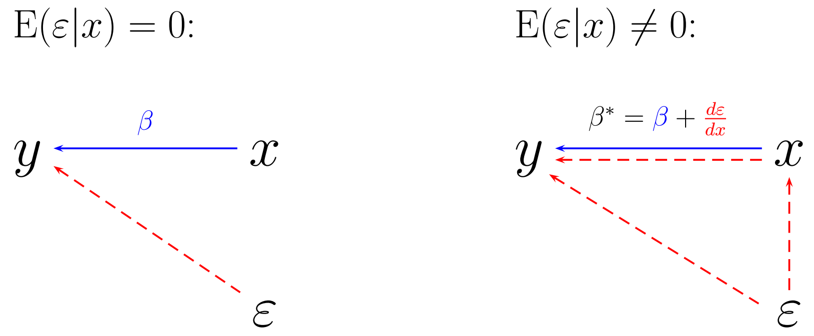


Abbildung 10.5: Wenn $\mathrm{E}(\varepsilon_i|x_1,\ldots,x_n) = 0$ (links) ist der OLS Schätzer unverzerrt, β misst den Effekt von x auf y. Falls $\mathrm{E}(\varepsilon_i|x_1,\ldots,x_n) \neq 0$ (rechts) ist der OLS Schätzer verzerrt; der messbare Parameter β^* misst den direkten Effekt β *und* indirekten Effekt $d\varepsilon/dx$ gemeinsam, der interessierende Parameter β ist in diesem Fall *nicht identifiziert*.

ε gewissermaßen 'verschmutzt'. Deshalb werden fehlende relevante Variablen in der englischen Literatur häufig als ***'confounder'*** bezeichnet.

Wenn ein Regressor x mit ε korreliert ist können wir uns vorstellen, ε ist eine Funktion von x, das heißt $\varepsilon = \varepsilon(x)$. In diesem Fall misst β_2 nicht mehr den marginalen Effekt von x, da

$$y = \beta_1 + \beta_2 x + \varepsilon(x)$$

In diesem Fall gibt es einen direkten und indirekten Zusammenhang zwischen x und y

$$\frac{dy}{dx} = \beta_2 + \frac{d\varepsilon}{dx} := \beta_2^*$$

Der OLS Schätzer misst den gemeinsamen Einfluss $\beta_2^* := \beta_2 + \frac{d\varepsilon}{dx}$ anstelle des meist interessierenden marginalen Effekts β_2, der OLS Schätzer ist systematisch verzerrt. Wie schon erwähnt kann dies auch als Identifikationsproblem interpretieren, mit einem endogenen Regressor x können wir nur den Gesamteffekt $\beta_2 + \frac{d\varepsilon}{dx}$ identifizieren, nicht aber den einzelnen interessierenden Effekt β_2.

Im Beispiel 'fehlender relevanter Variablen' war der datengenerierende Prozess (DGP) $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$, und wir interessieren uns für β_2. Wenn x_3 nicht beobachtet werden kann, aber $x_{i3} = \delta_1 + \delta_2 x_{i2} + \upsilon_i$ ist, erhalten wir durch einsetzen

$$y_i = \beta_1 + \beta_2 x_{i2} + \underbrace{[\beta_3(\delta_1 + \delta_2 x_{i2} + \upsilon_i) + \varepsilon_i]}_{\varepsilon^*}$$

und der marginale Effekt ist

$$\frac{\partial y}{\partial x_2} = \beta_2 + \frac{\partial \varepsilon^*}{\partial x_2} = \beta_2 + \beta_3 \delta_2$$

Wann immer der Regressor x mit dem Störterm ε korreliert ist gibt es *ohne zusätzlicher Information* keine Möglichkeit den Koeffizienten β_2 unverzerrt zu schätzen! Da diese Korrelation zwischen x und ε in der Regel auch mit zunehmender Stichprobengröße nicht verschwinden wird, ist der OLS Schätzer im Fall eines endogenen Regressors auch nicht konsistent.

Übungsbeispiel: Ein bekanntes Problem der Arbeitsmarktökonomik ist, dass die Stundenlöhne (StdL) nicht nur von der Bildung abhängen, sondern auch von unbeobachtbaren Fähigkeiten (F). Angenommen der datengenerierende Prozess wird durch

$$\log(\text{StdL}) = \beta_1 + \beta_2 \text{Bildg} + \beta_3 \text{F} + \varepsilon$$

beschrieben, aber sie haben keinen Indikator für die Fähigkeiten und schätzen deshalb

$$\log(\text{StdL}) = \widehat{\beta}_1^* + \widehat{\beta}_2^* \text{Bildg} + \hat{\varepsilon}^*$$

Mit welchem Bias würden Sei rechnen? Wovon hängt die Größe des Bias ab?

□

Zusammenfassend können wir festhalten, dass im Fall von *'Omitted Variables'*

- die OLS-Schätzer für die Koeffizienten verzerrt (*biased*) sind, wenn die inkludierten Variablen mit den nicht berücksichtigten Variablen korreliert sind;
- die OLS-Schätzer für die Koeffizienten auch *nicht* konsistent sind.
- Die Schätzer für die Standardabweichungen der Koeffizienten sind ebenfalls verzerrt, d.h. die üblichen Teststatistiken sind ungültig!
- Fehlende Variablen führen häufig zu Autokorrelation in den Residuen.

Hinweis: Während 'fehlende relevante Variablen' ziemlich verheerende Folgen für die Schätzung der Koeffizienten haben kann sind die Auswirkungen häufig weniger dramatisch, wenn das Ziel eine *Prognose* von $\widehat{y}$ ist. Wenn die fehlende Variable eine lineare Funktion der berücksichtigten Variablen ist sind die Vorhersagen sogar unverzerrt (Greene, 2003, 30).

Die Intuition ist einfach: angenommen der DGP sei $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$, aber wir schätzen irrtümlich das 'kurze' Modell $y_i = \widehat{\beta}_1^* + \widehat{\beta}_2^* x_{i2} + \hat{\varepsilon}_i^*$, und die fehlende Variable x_3 sei eine lineare Funktion von x_2, d.h. $x_{i3} = \delta_1 + \delta_2 x_{i2} + \upsilon_i$ mit $\text{E}(x_{i3}) = \delta_1 + \delta_2 x_{i2}$.

Wenn wir in das korrekte 'lange' Model einsetzen erhalten wir

$$\begin{aligned}
y_i &= \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i \\
&= \beta_1 + \beta_2 x_{i2} + \beta_3(\delta_1 + \delta_2 x_{i2} + \upsilon_i) + \varepsilon_i \\
&= \underbrace{\beta_1 + \beta_3\delta_1}_{\beta_1^*} + \underbrace{(\beta_2 + \beta_3\delta_2)}_{\beta_2^*} x_{i2} + \underbrace{\beta_3\upsilon_i + \varepsilon_i}_{\varepsilon_i^*} \\
&= \beta_1^* + \beta_2^* x_{i2} + \varepsilon_i^*
\end{aligned}$$

Wenn wir diese fehlspezifizierte Gleichung schätzen erhalten wir $y_i = \widehat{\beta}_1^* + \widehat{\beta}_2^* x_{i2} + \hat{\varepsilon}_i^*$ mit

$$\text{E}(y_i) = \text{E}(\beta_1^* + \beta_2^* x_{i2} + \varepsilon_i^*) = (\beta_1 + \beta_3\delta_1) + (\beta_2 + \beta_3\delta_2) x_{i2}$$

Die Koeffizienten des 'kurzen' Modells haben zwar keine kausale Interpretation, aber für die Vorhersage 'korrigieren' sie gewissermaßen für die fehlende Variable x_3. Natürlich ist die Vorhersage mit dem 'kurzen' Modell nicht effizient, wie man einfach am Störterm $\varepsilon_i^* = \beta_3\upsilon_i + \varepsilon_i$ erkennt.

Exkurs: Irrtümliche Berücksichtigung nicht relevanter Variablen (*Redundant Variables*)

Natürlich kann uns auch das Gegenteil passieren, dass wir irrtümlich Variablen berücksichtigen, die in Wahrheit keinen Einfluss auf y haben.

Dieser Fall hat weniger dramatische Auswirkungen, da dies zu *keiner* Korrelation mit dem Störterm führt, deshalb führt dies *nicht* zu Endogenität.

Dies kann wieder einfach gezeigt werden: das 'wahre Modell' (in Abweichungsform) sei

$$\ddot{y}_i = \beta_2 \ddot{x}_{i2} + \varepsilon_i$$

aber wir schätzen ein falsches ('langes') Modell

$$\ddot{y}_i = \beta_2^* \ddot{x}_{i2} + \beta_3^* \ddot{x}_{i3} + \varepsilon_i^*$$

d.h. $\ddot{x}_3$ wurde irrtümlich als erklärende Variable berücksichtigt, und die im 'wahren Modell' geltende Restriktion $\beta_3^* = 0$ wurde fälschlich nicht berücksichtigt.

Der OLS Schätzer für zwei erklärende Variablen ist in Summenschreibweise

$$\widehat{\beta}_2^* = \frac{(\sum \ddot{x}_{i3}^2)(\sum \ddot{x}_{i2}\ddot{y}_i) - (\sum \ddot{x}_{i2}\ddot{x}_{i3})(\sum \ddot{x}_{i3}\ddot{y}_i)}{(\sum \ddot{x}_{i2}^2)(\sum \ddot{x}_{i3}^2) - (\sum \ddot{x}_{i2}\ddot{x}_{i3})^2}.$$

Einsetzen von $\ddot{y}_i = \beta_2 \ddot{x}_{i2} + \varepsilon_i$ und lösen gibt:

$$\widehat{\beta}_2^* = \beta_2 + \frac{(\sum \ddot{x}_{i3}^2)(\sum \ddot{x}_{i2}\varepsilon_i) - (\sum \ddot{x}_{i2}\ddot{x}_{i3})(\sum \ddot{x}_{i3}\varepsilon_i)}{(\sum \ddot{x}_{i2}^2)(\sum \ddot{x}_{i3}^2) - (\sum \ddot{x}_{i2}\ddot{x}_{i3})^2}.$$

Wenn x_2 und x_3 mit ε_i unkorreliert sind (z.B. wenn x_2 und x_3 deterministisch sind) folgt für den Erwartungswert:

$$\mathrm{E}(\widehat{\beta}_2^*) = \beta_2$$

Bei fälschlicher Berücksichtigung einer irrelevanten Variablen bleibt die OLS-Schätzung also erwartungstreu und konsistent! Man kann jedoch zeigen, dass sie nicht varianzminimal, d.h. nicht effizient ist.

Zusammenfassend können wir festhalten, dass im Fall der Berücksichtigung irrelevanter Variablen (*'redundant variables'*):

- die OLS-Schätzer für die Koeffizienten erwartungstreu und konsistent bleiben,
- der OLS-Schätzer für die Koeffizienten *nicht* effizient sind, da sie nicht varianzminimal sind, und
- die Schätzer für die Standardabweichungen der Koeffizienten unverzerrt sind, d.h. die üblichen Teststatistiken bleiben gültig, aber die Schätzungen werden 'unscharf'.

10.4 Die Methode der Zweistufigen Kleinsten Quadrate

Die OLS Methode beruht auf einer orthogonalen Zerlegung von y in eine systematische – durch die x Variablen erklärte – Komponente, und in eine stochastische nicht-systematische Komponente, die Residuen $\hat{\varepsilon}$. Der OLS Schätzer stellt sicher, dass die Residuen mit den x Variablen *in der Stichprobe* unkorreliert sind (orthogonal), aber dies muss nicht für die Störterme der Grundgesamtheit ε gelten.

Endogenität im ökonometrischen Sinne ist definiert als eine Abhängigkeit zwischen Störtermen und Regressoren, d.h. $\mathrm{E}(\boldsymbol{\varepsilon}|\boldsymbol{X}) \neq \mathbf{0}$.

Kehren wir nochmals zurück zur Lernzeit und dem Prüfungserfolg. Wenn wir die Fähigkeiten nicht beobachten können führt dies zu verzerrten Schätzungen, die Störterme des fehlspezifizierten Modells $\varepsilon_i^* = \beta_3 F_i + \hat{\varepsilon}_i$ sind mit der Lernzeit korreliert.

Wenn wir aber in einem (hypothetischen) Experiment die Lernzeit durch eine Intervention (randomisiert) beeinflussen könnten, dann könnten wir dies nützen um den kausalen Effekt β_2 messen. Dies wird allerdings nur in den seltensten Fällen möglich sein.

Aber etwas ähnliches könnte gelingen, wenn wir eine Variable hätten, die wie eine Intervention *nur* die – durch die Fehlspezifikation endogene – Variable Lernzeit beeinflusst, und *nur* über die Lernzeit auf die Punktezahl wirkt.

Eine solche Variable wird **Instrumentvariable** genannt. Stellen wir uns vor, das Wetter (W) am Vorabend würde sich *nur* auf die Lernzeit auswirken, aber keine andere direkte oder indirekte Wirkung auf den Prüfungserfolg (Pkt) haben.

Das Wetter würde dann etwas ähnliches bewerkstelligen wie ein Experimentator durch seine Intervention, eine isolierte Beeinflussung der interessierenden Variable.

Intuitiv können wir uns vorstellen, dass in der Lernzeit indirekt die unterschiedlichen Fähigkeiten zum Ausdruck kommen, sie enthält deshalb gewissermaßen 'zwei Arten von Streuung', eine *erwünschte Streuung*, die zur Messung der systematischen Komponente von verwendet werden kann, und eine *unerwünschte Streuung*, die durch die Korrelation mit den Störtermen ε (d.h. der nicht-systematischen Komponente von y) zustande kommt, siehe Abbildung 10.5.

Diese intuitive Idee der Zerlegung der Streuung von x in eine 'gute' und erwünschte Streuung und in eine unerwünschte problematische Streuung führt unmittelbar zur Idee der Methode der zweistufigen kleinsten Quadrate (*'two stage least squares'* bzw. 2SLS).

Wenn das Wetter tatsächlich nur die Lernzeit beeinflusst und sonst keinen systematischen Einfluss hat können wir der Wetter nützen, um die Lernzeit mit Hilfe von OLS orthogonal in zwei Teile zu zerlegen.

Wir schätzen

$$\mathrm{LZ} = \hat{\alpha}_1 + \hat{\alpha}_2 \mathrm{W} + \hat{\upsilon} \quad \longrightarrow \quad \widehat{\mathrm{LZ}}$$

Der gefittete Wert $\widehat{\mathrm{LZ}}$ hängt nur vom Wetter ab und sollte deshalb nicht mit den Störtermen (und den darin enthaltenen Fähigkeiten) korreliert sein.

Diese gefitteten Werte können in einem zweiten Schritt verwendet werden, um den kausalen Effekt der Lernzeit auf die Punktezahl konsistent zu schätzen

$$\text{Pkt} = \widehat{\beta}_1 + \widehat{\beta}_2 \,\widehat{\text{LZ}} + \hat{\varepsilon}$$

In unserem Beispiel

1. In der ersten Stufe wird der (durch Fehlspezifikation) endogene Regressor auf die Instrumentvariable regressiert

$$\begin{array}{rcccc} \text{LZ} & = & 9.389 & - & 0.82\ \text{W} \\ & & (0.568)^{***} & & (0.178)^{***} \end{array}$$

$$R^2 = 0.268, \qquad n = 60$$

2. In der zweiten Stufe wird anstelle der tatsächlichen Lernzeit nur die durch das exogene Wetter erklärte Lernzeit als regressor verwendet

$$\text{Pkt} \;=\; 93.407 \;+\; 7.351\ \widehat{\text{LZ}}$$

   (da, wie wir später sehen werden, die Standardfehler der zweiten Stufe durch dieses Verfahren nicht korrekt berechnet werden, und auch das R^2 nicht die übliche Interpretation hat, werden sie hier nicht angegeben.)

Man kann zeigen, dass dieser Schätzer konsistent ist, allerdings ist auch er in kleinen Stichproben meist verzerrt.

In diesem Beispiel erhalten wir einen Punktschätzer von 7.35, d.h. aufgrund dieser Schätzung würden wir erwarten, dass eine zusätzliche Lernstunde im Durchschnitt 7.35 zusätzliche Punkte einbringt. Dies ist zwar immer noch eine ziemlich ungenaue Schätzung (der wahre Wert war 5), aber immerhin genauer als die Schätzung der kurzen Modells $\text{Pkt} = 157.404 - 1.701\,\text{LZ} + \hat{\varepsilon}$.

Damit diese Methode der Zweistufigen Kleinsten Quadrate konsistente Schätzfunktionen liefert müssen die **'Instrumentvariablen'** (z) die beiden folgenden Anforderungen erfüllen

1. **Relevanz:** $\text{cov}(z, x) \neq 0$, und
2. **Exogenität:** $\text{cov}(z, \varepsilon) = 0$

Abbildung 10.6 (rechtes Panel) zeigt das Grundprinzip einer 2SLS Schätzung. In der ersten Stufe wird der endogene Regressor x in zwei Teile zerlegt, einen exogenen Teil $\hat{x}$, der durch die Instrumentvariable erklärt wird, und in einen endogenen Teil υ_x. Die Streuung von $\hat{x}$ ist unproblematisch, da sie ausschließlich durch die per Definition exogene Instrumentvariable erklärt wird. Die problematische Streuung von υ_x wird verworfen, vgl Abbildung 10.6.

In der zweiten Stufe wird das exogene (gefittete) $\hat{x}$ als Regressor verwendet.

Wir werden später sehen, dass – wenn die Annahmen für Instrumentvariablen (*Relevanz* und *Exogenität*) erfüllt sind – Instrumentvariablenschätzer konsistent sind.

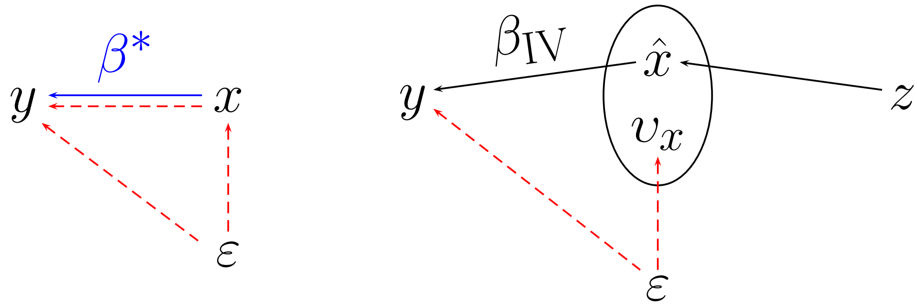


Abbildung 10.6: Falls $\mathrm{E}(\varepsilon|\boldsymbol{X}) \neq 0$ ist der OLS Schätzer β^* verzerrt, da der Einfluss von x auf y durch die Korrelation mit dem Störterm 'verschmutzt' wird (linkes Panel). Bei einer IV Schätzung wird x gewissermaßen in einen exogenen Teil $\hat{x}$ – der durch die Instrumentvariable erklärt wird – und einen endogenen Teil υ_x zerlegt. Die IV Schätzer verwenden nur die exogene Streuung $\hat{x}$ (rechtes Panel).

Die Überprüfung der Relevanz ist relativ einfach, wir müssen nur die Schätzung der ersten Stufe überprüfen; die Instrumentvariable(n) sollten einen möglichst großen Erklärungsbeitrag leisten.

Viel schwieriger ist die Exogenität zu überprüfen. Es ist wichtig zu erkennen, dass ökonometrische Exogenität eine deutlich strengere Annahme ist als das, was man z.B. in der Mikroökonomik darunter versteht.

Das Wetter ist aus einer mikroökonomischen Perspektive sicher exogen, aber wenn unsere Prüflinge wetterfühlig sind kann dies zu einer Korrelation mit den nicht berücksichtigten Fähigkeiten führen, und damit zu *ökonometrischer Endogenität*!!!

Die tatsächliche Berechnung der Schätzer erfolgt in der Praxis allerdings nicht nach diesem zweistufigen Verfahren, sondern mit Hilfe eine allgemeinerer Instrumentvariablen Schätzer, aber dieses zweistufigen Verfahren führt in diesem einfachen Fall numerisch zu den gleichen Werten und liefert eine bessere Intuition.

10.5 Instrumentvariablen Schätzer (IV)

> "Two things distinguish the discipline of econometrics from the older sister field of statistics. One is the lack of shyness about causality. Causal inference has always been the name of the game in applied econometrics. [...]
>
> The second thing that distinguishes us from most statisticians – and indeed from most other social scientists – is an arsenal of statistical tools that grew out of early econometric research on the problem of how to estimate the parameters in a system of linear simultaneous equations. The most powerful weapon in this arsenal is the method of instrumental variables (IV), the subject of this chapter." (Angrist and Pischke, 2008, p. 113)

Es gibt eine ganze Reihe von *Instrumentvariablen Schätzern* (IV Schätzer), wir werden hier nur eine einfache Variante nach der 'Methode der Momente' herleiten. Diese liefern numerisch die gleichen Werte wie die Methode der zweistufigen kleinste Quadrate (2SLS), sind aber einfacher anzuwenden und können leichter verallgemeinert werden.

Da Instrumentvariablen Schätzer *keine* OLS Schätzer sind ist das Gauss Markov Theorem nicht anwendbar, deshalb garantiert uns nichts die Unverzerrtheit und Effizienz der IV Schätzer. Tatsächlich sind IV Schätzer in kleinen Stichproben meist verzerrt und konvergieren nur mit zunehmender Stichprobengröße zu den wahren Werten, d.h. sie sind konsistent.

Technisch werden IV Schätzer meist mit Hilfe der 'Methode der Momente' hergeleitet, die wir im folgenden kurz vorstellen werden. 'Methode der Momente' Schätzer sind zwar nicht immer unverzerrt und effizient, aber sie sind generell konsistent.

10.5.1 Instrumentvariablen als Methode der Momente Schätzer

Wir haben bisher erst eine Methode zur Schätzung eines unbekannten Parametervektors kennengelernt, nämlich die *Methode der Kleinsten Quadrate* (OLS). Die historisch älteste Methode ist allerdings die *'Methode der Momente'*, die immer konsistente Schätzer liefert, die aber – im Unterschied zu Maximum Likelihood Schätzern – nicht immer asymptotisch effizient sind. Die Grundidee der Methode der Momente Schätzer besteht darin, dass die Momente der Stichprobe (z.B. Mittelwert, Varianz, ...) als Schätzer für die Momente der Grundgesamtheit herangezogen werden.

Wir zeigen dies für das einfache Regressionsmodell. In der Grundgesamtheit bestehe folgender Zusammenhang

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$$

Wir nehmen für die Momente der Grundgesamtheit an, dass

- der Erwartungswert der Störterme der Grundgesamtheit ist Null

$$\mathrm{E}(\varepsilon_i) = 0$$

- die Kovarianz zwischen den Störtermen der Grundgesamtheit ε_i und dem möglicherweise stochastischen x_i ist ebenfalls Null

$$\mathrm{E}(x_i \varepsilon_i) = 0$$

Die entsprechenden Bedingungen *für die Stichprobe* sind

$$\frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_i = 0 \qquad \text{und} \qquad \frac{1}{n}\sum_{i=1}^{n} x_i \hat{\varepsilon}_i = 0$$

wobei $\hat{\varepsilon}_i$ die Stichprobenresiduen sind.

Durch Einsetzen der Stichprobenresiduen $\hat{\varepsilon}_i = y_i - \widehat{\beta}_1 - \widehat{\beta}_2 x_i$ in die obigen Bedingungen erhalten wir die beiden Normalgleichungen

$$\begin{aligned} \sum y_i &= n\widehat{\beta}_1 + \widehat{\beta}_2 \sum x_i \\ \sum y_i x_i &= \widehat{\beta}_1 \sum x_i + \widehat{\beta}_2 \sum x_i^2 \end{aligned}$$

wobei $\widehat{\beta}_1$ und $\widehat{\beta}_2$ die Methode der Momente Schätzer für β_1 und β_2 sind.

Dies sind aber exakt die gleichen Normalgleichungen, die wir auch für OLS erhalten haben, deshalb liefert die Methode der Momente in diesem einfachen Fall exakt die gleichen Schätzer wie die Methode der kleinsten Quadrate oder – wie wir in einem späteren Kapitel zeigen werden – auch die Maximum Likelihood Methode.

In diesem Fall einfachen führen also alle drei bekannten Schätzmethoden, d.h. OLS, Maximum Likelihood und die Methode der Momente, zum gleichen Ergebnis, aber dies gilt natürlich nicht immer. In manchen komplexen Situationen bietet die Methode der Momente eine relativ einfache Möglichkeit um zu einem konsistenten Schätzer zu gelangen, z.B. bei der Herleitung des Instrumentenvariablen-Schätzers.

Mit dem Methode der Momente Schätzer haben wir die Bedingungen

$$\mathrm{E}(\varepsilon_i) = 0 \qquad \text{und} \qquad \mathrm{E}(x_i \varepsilon_i) = 0$$

angenommen um einen konsistenten Schätzer zu erhalten, wobei x auch stochastisch sein kann. Wenn aber zum Beispiel eine fehlerhaft gemessene Variable x^* mit ε korreliert ist kann *nicht* angenommen werden, dass $\mathrm{E}(x_i^* \varepsilon_i) = 0$.

Aber angenommen, wir hätten wieder eine Variable z zur Verfügung, die nicht mit den ε korreliert ist, also $\mathrm{E}(z_i, \varepsilon_i) = 0$, die aber mit x_i korreliert ist ($\mathrm{cov}(z_i, x_i) \neq 0$), kann die Methode der Momente herangezogen werden um zumindest einen konsistenten Schätzer zu finden.

Die entsprechenden Stichproben Momente sind in diesem Fall

$$\frac{1}{n} \sum \hat{\varepsilon}_i = 0 \qquad \text{und} \qquad \frac{1}{n} \sum (z_i \hat{\varepsilon}_i) = 0$$

Eine solche **Instrumentvariable** (z) muss – wie wir bereits im letzten Abschnitt gesehen haben – zwei wichtige Bedingungen erfüllen:

1. die Instrumentvariable z sollte möglichst hoch mit dem x korreliert sein, d.h. sie muss *relevant* sein
$$\mathrm{cov}(z, x) \neq 0$$
2. die Instrumentvariable z darf nicht mit dem Störterm ε der Grundgesamtheit korreliert sein, d.h. sie muss *exogen* sein
$$\mathrm{cov}(z, \varepsilon) = 0$$

Man beachte, dass die zweite Bedingung, $\mathrm{cov}(z, \varepsilon) = 0$, nicht einfach getestet werden kann, da die Störterme der Grundgesamtheit unbekannt sind![3] Man kann auch nicht

[3]Ein Test auf Endogenität des Regressors ist der *Hausman Test*, der später vorgestellt wird.

die Stichprobenresiduen $\hat{\varepsilon}$ für einen solchen Test nützen, da diese per Konstruktion mit x unkorreliert sind.

Hingegen kann die erste Bedingung $\operatorname{cov}(z, x) \neq 0$ sehr einfach getestet werden, indem man eine Regression $x_i = a_0 + a_1 z_i + \nu_i$ rechnet und mit dem üblichen t-Test überprüft, ob a_1 signifikant von Null verschieden ist (bzw. mit einem F-Test im multivariaten Fall).

Die Instrumentvariablen Schätzer $\widehat{\beta}_{k,\mathrm{IV}}$ erhalten wir, indem wir die IV-Residuen $\hat{\varepsilon}_i = y_i - \widehat{\beta}_{1,\mathrm{IV}} - \widehat{\beta}_{2,\mathrm{IV}} x_i$ in die Momentbedingungen

$$\frac{1}{n}\sum \hat{\varepsilon}_i = 0 \qquad \text{und} \qquad \frac{1}{n}\sum (z_i \hat{\varepsilon}_i) = 0$$

einsetzen

$$\begin{aligned}
\sum \left(y_i - \widehat{\beta}_{1,\mathrm{IV}} - \widehat{\beta}_{2,\mathrm{IV}} x_i\right) &= 0 \\
\sum z_t \left(y_i - \widehat{\beta}_{1,\mathrm{IV}} - \widehat{\beta}_{2,\mathrm{IV}} x_i\right) &= 0
\end{aligned}$$

Daraus können wieder die Normalgleichungen hergeleitet werden

$$\begin{aligned}
\sum y_i &= n\widehat{\beta}_{1,\mathrm{IV}} + \widehat{\beta}_{2,\mathrm{IV}} \sum x_i \\
\sum y_i z_i &= \widehat{\beta}_{1,\mathrm{IV}} \sum z_i + \widehat{\beta}_{2,\mathrm{IV}} \sum x_i z_i
\end{aligned}$$

Diese Gleichungen können wieder wie üblich gelöst werden, indem wir die erste Gleichung mit $\sum z_i$ und die zweite Gleichung mit n multiplizieren sowie die erste Gleichung von der zweiten Gleichung subtrahieren,

$$\widehat{\beta}_{2,\mathrm{IV}} = \frac{n\sum y_i z_i - \sum y_i \sum z_i}{n\sum x_i z_i - \sum x_i \sum z_i}$$

Dies kann ähnlich wie früher vereinfacht werden und gibt die folgenden Instrumentvariablen Schätzer für den bivariaten Fall

$$\begin{aligned}
\widehat{\beta}_{2,\mathrm{IV}} &= \frac{\sum (y_i - \bar{y})(z_i - \bar{z})}{\sum (x_i - \bar{x})(z_i - \bar{z})} = \frac{\sum \ddot{y}_i \ddot{z}_i}{\sum \ddot{x}_i \ddot{z}_i} = \frac{\operatorname{cov}(y, z)}{\operatorname{cov}(x, z)} \\
\widehat{\beta}_{1,\mathrm{IV}} &= \bar{y} - \widehat{\beta}_{2,\mathrm{IV}} \bar{x}
\end{aligned}$$

wobei $\bar{x} := 1/n \sum x_i$ wieder den Mittelwert von x bezeichnet und $\ddot{x}_i := x_i - \bar{x}$ die Abweichungen vom Mittelwert sind, bzw. $\ddot{z}_i := z_i - \bar{z}$.

Man beachte, dass für den IV Schätzer nicht einfach x durch z 'ersetzt' wird, im Nenner des Schätzers für $\widehat{\beta}_{2,\mathrm{IV}}$ steht nämlich weder $\operatorname{var}(x)$ noch $\operatorname{var}(z)$, sondern $\operatorname{cov}(x, z)$!

Intuition:* Marginaler Effekt und Kettenregel
(Cameron, Trivedi, Microeconometrics using Stata, p. 173)

Wenn x ein endogener Regressor und z eine Instrumentvariable ist gilt

$$\widehat{\beta}_{\text{IV}} = \frac{dy}{dz}\frac{dz}{dx}$$

Warum?

$$\widehat{\beta}_{\text{IV}} = \frac{\sum_i \ddot{y}_i \ddot{z}_i}{\sum_i \ddot{x}_i \ddot{z}_i} = \frac{\frac{\sum_i \ddot{y}_i \ddot{z}_i}{\sum_i \ddot{z}_i^2}}{\frac{\sum_i \ddot{x}_i \ddot{z}_i}{\sum_i \ddot{z}_i^2}}$$

Sei

$$\hat{\ddot{y}}_i = \hat{\alpha}\, \ddot{z}_i \qquad \text{und} \qquad \hat{\ddot{x}}_i = \hat{\delta}\, \ddot{z}_i$$

mit

$$\hat{\alpha} = \frac{\sum_i \ddot{y}_i \ddot{z}_i}{\sum_i \ddot{z}_i^2} = \frac{d\,\hat{\ddot{y}}}{d\,\ddot{z}} \qquad \text{und} \qquad \hat{\delta} = \frac{\sum_i \ddot{x}_i \ddot{z}_i}{\sum_i \ddot{z}_i^2} = \frac{d\,\hat{\ddot{x}}}{d\,\ddot{z}}$$

Deshalb

$$\widehat{\beta}_{\text{IV}} = \frac{\hat{\alpha}}{\hat{\delta}} = \frac{\frac{d\,\hat{\ddot{y}}}{d\,\ddot{z}}}{\frac{d\,\hat{\ddot{x}}}{d\,\ddot{z}}} = \frac{d\,\hat{\ddot{y}}}{d\,\ddot{z}}\frac{d\,\ddot{z}}{d\,\hat{\ddot{x}}}$$

■

Um die Konsistenz dieses Schätzers zu zeigen verwenden wir wieder die Variablen in Abweichungsform vom Mittelwert und setzen für $\ddot{y}$ den wahren Wert $\beta_2 \ddot{x}_i + \varepsilon_i$ ein. Wir erhalten

$$\widehat{\beta}_{2,\text{IV}} = \beta_2 + \frac{\sum \ddot{z}_i \varepsilon_i}{\sum \ddot{x}_i \ddot{z}_i}$$

Wenn $\ddot{x}_i$ stochastisch ist kann der Erwartungswert nicht einfach ermittelt werden, da wir hier einen Quotienten zweier Zufallsvariablen haben, aber der plim kann einfach berechnet werden

$$\text{plim}(\widehat{\beta}_{2,\text{IV}}) = \beta_2 + \frac{\text{plim}\left[\frac{1}{n}\sum(\ddot{z}_i \varepsilon_i)\right]}{\text{plim}\left[\frac{1}{n}\sum \ddot{x}_i \ddot{z}_i\right]}$$

Der Zähler des letzten Ausdrucks der rechten Seite ist Null, da annahmegemäß $\text{cov}(\ddot{z}_i \varepsilon_i) = 0$. Der Nenner ist der plim der Stichprobenkovarianz zwischen $\ddot{x}_i$ und $\ddot{z}_i$ und ist deshalb ein konsistenter Schätzer für die Kovarianz zwischen $\ddot{x}_i$ und $\ddot{z}_i$ der Grundgesamtheit $\sigma_{\ddot{x}z}$. Wenn $\sigma_{\ddot{x}\ddot{z}} \neq 0$ folgt also

$$\widehat{\beta}_{2,\text{IV}} = \beta_2 + \frac{0}{\sigma_{\ddot{x}\ddot{z}}} = \beta_2$$

Man beachte, dass selbst eine sehr geringe Korrelation zwischen z und ε zu großen Fehlern führen kann, wenn die Korrelation zwischen z und x ebenfalls sehr klein ist. Deshalb sollte man die Korrelation zwischen x und z *immer* mit einer Hilfsregression überprüfen!

Die Konsistenz von $\widehat{\beta}_{1,\text{IV}}$ kann ähnlich gezeigt werden.

Da unterschiedliche Instrumente gewählt werden können, und jedes Instrument zu einem anderen konsistenten Schätzer führt, kann nicht davon ausgegangen werden, dass Instrumentvariablen Schätzer asymptotisch effizient sind.

In großen Stichproben gilt allerdings, dass

$$\widehat{\beta}_{2,\mathrm{IV}} \stackrel{a}{\sim} \mathrm{N}\left(\beta_2, \mathrm{var}(\widehat{\beta}_{2,\mathrm{IV}})\right)$$

mit der geschätzten Varianz

$$\widehat{\mathrm{var}}\left(\widehat{\beta}_{2,\mathrm{IV}}\right) = \frac{\hat{\sigma}^2 \sum (z_i - \bar{z})^2}{\left[\sum (x_i - \bar{x})(z_i - \bar{z})\right]^2} = \frac{\hat{\sigma}^2 \sum \ddot{z}_i^2}{\left[\sum \ddot{x}_i \ddot{z}_i\right]^2}$$

Einen asymptotischen Schätzer $\hat{\sigma}^2$ für σ^2 erhalten wir aus

$$\hat{\sigma}^2 = \frac{1}{n} \sum \hat{\varepsilon}_i^2$$

wobei $\hat{\varepsilon}_i = y_i - \widehat{\beta}_{1,\mathrm{IV}} - \widehat{\beta}_{2,\mathrm{IV}} x_i$.

Daraus wird offensichtlich, dass die Varianz von $\widehat{\beta}_{2,\mathrm{IV}}$ umso kleiner ist, je größer die Korrelation zwischen x_i und z_i ist.

Für das frühere Beispiel mit der Lernzeit erhalten wir eine IV Schätzung mit konsistenten Standardfehlern

```
insheet using "https://www.uibk.ac.at/econometrics/data/iv_bsp_lz.csv"
ivregress 2sls pkt (lz = w)
```

$$\begin{array}{rcccc} \mathrm{Pkt} & = & 93.407 & + & 7.351\ \mathrm{LZ} \\ & & (19.917)^{***} & & (2.777)^{**} \end{array}$$

$$R^2 = -1.946, \qquad n = 60$$

Wie man sieht kann diese Methode zu einem negativen R^2 führen.

Instrumentvariablen dürfen nicht mit Proxy-Variablen verwechselt werden. Proxy Variablen sollten eine fehlende Variable gewissermaßen 'ersetzen', kommen also in der Regressionsgleichung direkt vor, während Instrumentvariablen nur 'indirekt' vorkommen (d.h. nicht als Regressor aufscheinen).

Die Instrumentvariablen z sollten möglichst hoch mit dem endogenen Regressor x korreliert sein, und *nur* über x auf y einwirken, aber sie dürfen *nicht* mit fehlenden Variablen korreliert sein (d.h. sie müssen unkorreliert mit ε sein)!

Proxy-Variablen sollten möglichst hoch mit den fehlenden Variablen korreliert sein, während Instrumentvariablen nicht mit den fehlenden Variablen korreliert sein dürfen.

Im Falle einer Lohngleichung $\log(\mathrm{Stdl} = \widehat{\beta}_1 + \widehat{\beta}_2 \mathrm{Bildg} + \widehat{\beta}_3 \mathrm{Fähigkeiten} + \hat{\varepsilon}$ könnte z.B. als Proxy-Variable für die unbeobachtbaren Fähigkeiten der gemessene Intelligenzquotient verwendet werden.

Eine Instrumentvariable sollte hingegen mit den unbeobachtbaren Fähigkeiten und dem Störterm unkorreliert sein, aber möglichst hoch mit der Variablen 'Bildg' korrelieren. Arbeitsmarktökonomen haben z.B. die Entfernung zwischen Heimat- und Studienort (Card 1995, *Aspects of Labour Market Behavior*) oder die Anzahl der

Geschwister als Instrumentvariable für 'Bildung' vorgeschlagen, da Jugendliche aus kinderreichen Familien typischerweise eine kürzere Schulbildung erhalten, aber (hoffentlich) nicht über weniger Fähigkeiten verfügen (vgl. Wooldridge 2000, Chapter 15.1).[4]

10.6 Instrumentvariablen Schätzer für multiple Regressionen

Die grundlegende Idee kann leicht verallgemeinert werden für multiple Regressionen mit mehreren Instrumenten.

Im multivariaten Fall muss *mindestens* für jeden *endogenen* Regressor eine Instrumentvariable zur Verfügung stehen.

Allgemeiner, die IV Schätzer sind

- *exakt identifiziert*, wenn die Anzahl der Instrumente *gleich* der Anzahl der endogenen Regressoren ist.
- *überidentifiziert*, wenn die Anzahl der Instrumente *größer* ist als die Anzahl der endogenen Regressoren.
- *unteridentifiziert*, wenn die Anzahl der Instrumente *kleiner* ist als die Anzahl der endogenen Regressoren.

Unteridentifizierte Systeme können nicht geschätzt werden!

10.6.1 IV-Schätzer für den multivariaten Fall bei exakter Identifikation

Für den den multivariaten Fall haben wir bereits gezeigt

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}$$

woraus folgt

$$\operatorname{plim}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} + \operatorname{plim}\left(\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right) \cdot \operatorname{plim}\left(\frac{1}{n}(\boldsymbol{X}'\boldsymbol{\varepsilon})\right)$$

Wenn wir wieder annehmen, dass $\operatorname{plim}(\boldsymbol{X}'\boldsymbol{X}/n) = \boldsymbol{\Sigma}_{XX}$ eine positiv definite Matrix mit vollem Rang ist und $\operatorname{plim}(\boldsymbol{X}'\boldsymbol{\varepsilon}/n) = \boldsymbol{\Sigma}_{X\boldsymbol{\varepsilon}} \neq 0$, dann ist

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \boldsymbol{\Sigma}_{XX}^{-1} \cdot \boldsymbol{\Sigma}_{X\varepsilon}$$

kein konsistenter Schätzer, da einer oder mehrere der Regressoren mit dem Störterm $\boldsymbol{\varepsilon}$ der Grundgesamtheit korreliert ist.

[4] Allerdings könnten Kinder aus kinderreichen Familien über mehr soziale Fähigkeiten verfügen. Würde dies die Qualität der Instrumentvariable 'Anzahl der Geschwister' beeinflussen?

Das Modell ist also $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ mit $\mathrm{var}(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{I}$, aber mit $\mathrm{plim}\, \boldsymbol{X}'\boldsymbol{\varepsilon}/n \neq 0$!

Die Methode der Instrumenten-Variablen liefert zumindest konsistente Schätzer, wann immer der Störterm $\boldsymbol{\varepsilon}$ mit den erklärenden Variablen $\boldsymbol{X}$ korreliert ist. Wir fassen die Instrumente in einer $n \times k$ Matrix $\boldsymbol{Z} = (\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_k)$ zusammen, wobei für jede erklärende Variable ein Instrument mit n Beobachtungen existieren muss.[5] Diese Matrix $\boldsymbol{Z}$ darf nicht singulär sein und muss folgende Eigenschaften erfüllen:

$$
\begin{aligned}
\mathrm{plim}\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{\varepsilon}\right) &= 0 \\
\mathrm{plim}\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{X}\right) &= \boldsymbol{\Sigma}_{ZX} \quad \text{existiert und ist nicht singulär.} \\
\mathrm{plim}\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{y}\right) &= \boldsymbol{\Sigma}_{Zy} \quad \text{existiert und ist nicht singulär}
\end{aligned}
$$

Die erste Bedingung garantiert, dass die Korrelation zwischen Störterm und Instrumenten–Variable *'in Wahrscheinlichkeit gegen Null konvergiert'*. Die zweite Annahme garantiert eine positive Korrelation zwischen den X und den Z. Die dritte Gleichung ist lediglich eine Definition und unproblematisch.

Wir prämultiplizieren die ursprüngliche Gleichung $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ mit der transponierten Instrument-Matrix $\boldsymbol{Z}$ und erhalten

$$
\boldsymbol{Z}'\boldsymbol{y} = \boldsymbol{Z}'\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}'\boldsymbol{\varepsilon}
$$

anschließend dividieren wir beide Seiten durch n und bilden das *probability limit*

$$
\begin{aligned}
\mathrm{plim}\,\frac{\boldsymbol{Z}'\boldsymbol{y}}{n} &= \mathrm{plim}\,\frac{\boldsymbol{Z}'\boldsymbol{X}}{n}\boldsymbol{\beta} + \mathrm{plim}\,\frac{\boldsymbol{Z}'\boldsymbol{\varepsilon}}{n} \\
\boldsymbol{\Sigma}_{Zy} &= \boldsymbol{\Sigma}_{ZX}\boldsymbol{\beta} + \boldsymbol{0}
\end{aligned}
$$

da $\mathrm{plim}(\boldsymbol{Z}'\boldsymbol{\varepsilon}/n) = 0$ können wir nach $\boldsymbol{\beta}$ lösen

$$
\boldsymbol{\beta} = \boldsymbol{\Sigma}_{ZX}^{-1}\boldsymbol{\Sigma}_{Zy}
$$

Auf diese Gleichung können wir die Methode der Momente anwenden und die Momente der Grundgesamtheit $\boldsymbol{\Sigma}_{yX}$ und $\boldsymbol{\Sigma}_{Zy}$ durch die aus der Stichprobe berechneten Momente $\boldsymbol{y}'\boldsymbol{X}/n$ und $\boldsymbol{Z}'\boldsymbol{X}/n$ ersetzen. Dies gibt uns den Instrumentvariablen-Schätzer

$$
\hat{\boldsymbol{\beta}}^{\mathrm{IV}} = \left(\frac{\boldsymbol{Z}'\boldsymbol{X}}{n}\right)^{-1}\frac{\boldsymbol{Z}'\boldsymbol{y}}{n} = (\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{Z}'\boldsymbol{y}
$$

Wie man einfach sehen kann resultiert der OLS–Schätzer als Spezialfall, wenn $\boldsymbol{Z} = \boldsymbol{X}$, d.h. wenn jede x Variable als Instrument für sich selbst dient. Dies ist erlaubt, wenn die x exogen sind. Wenn hingegen die Instrumente überhaupt nicht mit den ursprünglichen Variablen korreliert sind, d.h. wenn $\boldsymbol{Z}'\boldsymbol{X} = \boldsymbol{0}$, bricht dieser Ansatz zusammen. Aber auch wenn die Korrelation zwischen den x und z Variablen gering ist liefert dieser Ansatz sehr schlechte Resultate $(\boldsymbol{Z}'\boldsymbol{X})$!

[5] Allerdings kann (und muss) eine Variable x, die nicht mit dem Störterm korreliert ist, als ihr eigenes Instrument verwendet werden!

Um die Konsistenz zu prüfen gehen wir wie üblich vor. Wir wissen, dass

$$\begin{aligned}\hat{\boldsymbol{\beta}}^{\mathrm{IV}} &= (\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{Z}'(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}) \\ &= (\boldsymbol{Z}'\boldsymbol{X})^{-1}(\boldsymbol{Z}'\boldsymbol{X})\boldsymbol{\beta}+(\boldsymbol{Z}'\boldsymbol{X})^{-1}(Z'\boldsymbol{\varepsilon}) \\ &= \boldsymbol{\beta}+(\boldsymbol{Z}'\boldsymbol{X})^{-1}(\boldsymbol{Z}'\boldsymbol{\varepsilon})\end{aligned}$$

Daraus folgt

$$\begin{aligned}\operatorname{plim}(\hat{\boldsymbol{\beta}}^{\mathrm{IV}}) &= \boldsymbol{\beta}+\operatorname{plim}\left[(\boldsymbol{Z}'\boldsymbol{X})^{-1}(\boldsymbol{Z}'\boldsymbol{\varepsilon})\right] \\ &= \boldsymbol{\beta}+\operatorname{plim}\left[\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{\varepsilon}\right)\right] \\ &= \boldsymbol{\beta}+\boldsymbol{\Sigma}_{ZX}^{-1}\mathbf{0}=\boldsymbol{\beta}\end{aligned}$$

womit gezeigt wurde, dass der Instrumentvariablen Schätzer tatsächlich konsistent ist.

Als nächstes müssen wir die asymptotische Varianz– Kovarianzmatrix $\boldsymbol{V}$ für $\hat{\boldsymbol{\beta}}^{\mathrm{IV}}$ ermitteln. Unter Berücksichtigung von $\hat{\boldsymbol{\beta}}^{\mathrm{IV}}-\boldsymbol{\beta}=(\boldsymbol{Z}'\boldsymbol{X})^{-1}(\boldsymbol{Z}'\boldsymbol{\varepsilon})$ erhalten wir

$$(\hat{\boldsymbol{\beta}}^{\mathrm{IV}}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}^{\mathrm{IV}}-\boldsymbol{\beta})'=(\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{Z}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{Z}(\boldsymbol{X}'\boldsymbol{Z})^{-1}$$

und

$$\operatorname{var}(\hat{\boldsymbol{\beta}}^{\mathrm{IV}})=\left[\frac{1}{n}\operatorname{plim}\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\right]\left[\operatorname{plim}\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{Z}\right)\right]\left[\operatorname{plim}\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)^{-1}\right]$$

woraus unter Bezugnahme auf die obigen Annahmen über die Instrumente folgt

$$\operatorname{var}(\hat{\boldsymbol{\beta}}^{\mathrm{IV}})=\frac{1}{n}\sigma^2\boldsymbol{\Sigma}_{ZX}^{-1}\boldsymbol{\Sigma}_{ZZ}\boldsymbol{\Sigma}_{ZX}^{-1}$$

Praktisch erfolgt die Schätzung einer konsistenten Varianz–Kovarianzmatrix durch

$$\hat{\sigma}^2(\boldsymbol{Z}'\boldsymbol{X})^{-1}(\boldsymbol{Z}'\boldsymbol{Z})(\boldsymbol{X}'\boldsymbol{Z})^{-1}$$

wobei $\hat{\sigma}^2$ (ein konsistenter Schätzer für σ^2) folgendermaßen ermittelt wird:

$$\hat{\sigma}^2=\frac{1}{n-k}\left(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}^{\mathrm{IV}}\right)'\left(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}^{\mathrm{IV}}\right)=\frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{n-k}$$

wobei es keine große Rolle spielt, ob wir durch n oder $n-k$ dividieren, da es sich um einen asymptotischen Test handelt.

Man beachte, dass all dies nur anwendbar ist, wenn für jede potentiell endogene x Variable genau eine Instrumentvariable zur Verfügung steht, weil sonst $\boldsymbol{Z}'\boldsymbol{X}$ nicht definiert ist! Der GIVE Schätzer ist auch für überidentifizierte Systeme anwendbar.

10.6.2 Der Verallgemeinerte Instrumentvariablenschätzer (GIVE)

GIVE (*'Generalized Instrumental Variable Estimator'*) ist ein Instrumentvariablenschätzer für den Fall, dass die Anzahl der Instrumente größer ist als die Zahl der Regressoren.

Bisher haben wir stets angenommen, dass wir exakt gleich viele Instrumente wie Regressoren haben, d.h. dass die Instrumenten Matrix $\boldsymbol{Z}$ und die Regressor Matrix $\boldsymbol{X}$ beide die gleiche Dimension $(n \times k)$ haben, wobei exogene X als Instrumente für sich selbst verwendet werden.

Wenn die Anzahl der Instrumente l größer ist als die Anzahl der X-Variablen k kann man die ursprüngliche Gleichung $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$ mit $\boldsymbol{X'Z(Z'Z)^{-1}Z'}$ vormultiplizieren

$$\boldsymbol{X'Z(Z'Z)^{-1}Z'y} = \boldsymbol{X'Z(Z'Z)^{-1}Z'X\beta} + \boldsymbol{X'Z(Z'Z)^{-1}Z'\varepsilon}$$

Wenn der letzte Term in Wahrscheinlichkeit gegen Null konvergiert und wir mit der Matrix $[\boldsymbol{X'Z(Z'Z)^{-1}Z'}]^{-1}$ vormultiplizieren erhalten wir nach Anwendung der Methode der Momente einen allgemeineren Instrumentvariablen Schätzer

$$\hat{\boldsymbol{\beta}}^{\mathrm{IV}} = \left[\boldsymbol{X'Z(Z'Z)^{-1}Z'X}\right]^{-1} \boldsymbol{X'Z(Z'Z)^{-1}Z'y}$$

Wenn wir $\boldsymbol{P}_Z := \boldsymbol{Z(Z'Z)^{-1}Z'}$ definieren können wir dies einfacher anschreiben

$$\hat{\boldsymbol{\beta}}^{\mathrm{IV}} = [\boldsymbol{X'P}_Z\boldsymbol{X}]^{-1}\, \boldsymbol{X'P}_Z\boldsymbol{y}$$

Die entsprechende Varianz-Kovarianzmatrix der Parameter ist

$$\mathrm{var}(\hat{\boldsymbol{\beta}}^{\mathrm{IV}}) = \sigma^2(\boldsymbol{X'P}_Z\boldsymbol{X})^{-1}$$

wobei ein Schätzer $\hat{\sigma}^2$ für σ^2 wie üblich aus den Residuen berechnet werden kann (zeigen Sie, dass $\boldsymbol{P}_Z$ symmetrisch und idempotent ist).

Die Intuition dahinter wird klarer, wenn man beachtet, dass $\boldsymbol{P}_Z := \boldsymbol{Z(Z'Z)^{-1}Z'}$ eine Projektionsmatrix ist!

Wenn die die Matrix $\boldsymbol{X}$ mit der Projektionsmatrix $\boldsymbol{P}_Z$ vormultipliziert wird, werden die $\boldsymbol{X}$ in den Spaltenraum der Instrumente $\boldsymbol{Z}$ projiziert, d.h.

$$\widehat{\boldsymbol{X}} = \boldsymbol{P}_Z\boldsymbol{X}$$

Daraus folgt eine alternative Möglichkeit den den IV Schätzer herzuleiten. Man kann eine einfache Datentransformation vornehmen, indem man das Modell $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$ mit der Projektionsmatrix $\boldsymbol{P}_Z$ vormultipliziert

$$\underbrace{\boldsymbol{P}_Z\boldsymbol{y}}_{\widehat{\boldsymbol{y}}} = \underbrace{\boldsymbol{P}_Z\boldsymbol{X}}_{\widehat{\boldsymbol{X}}}\boldsymbol{\beta} + \underbrace{\boldsymbol{P}_Z\boldsymbol{\varepsilon}}_{\widehat{\boldsymbol{\varepsilon}}}$$

und für die derart transformierten Daten $\widehat{\boldsymbol{y}} = \widehat{\boldsymbol{X}}\boldsymbol{\beta} + \widehat{\boldsymbol{\varepsilon}}$ den OLS Schätzer anwendet. Der OLS Schätzer auf die derart transformierten Daten ist der IV Schätzer

$$\hat{\boldsymbol{\beta}}^{\mathrm{IV}} = (\widehat{\boldsymbol{X}}'\widehat{\boldsymbol{X}})^{-1}\widehat{\boldsymbol{X}}'\hat{\mathrm{y}}$$

weil

$$\begin{aligned}
[(\boldsymbol{P}_Z\boldsymbol{X})'\boldsymbol{P}_Z\boldsymbol{X}]^{-1}(\boldsymbol{P}_Z\boldsymbol{X})'\boldsymbol{P}_Z\boldsymbol{y} &= \\
[\boldsymbol{X}'\boldsymbol{P}_Z'\boldsymbol{P}_Z\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{P}_Z'\boldsymbol{P}_Z\boldsymbol{y} &= \\
[\boldsymbol{X}'\boldsymbol{P}_Z\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{P}_Z\boldsymbol{y} &= \\
\left[\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}\right]^{-1}\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{y} &= \hat{\boldsymbol{\beta}}^{\text{IV}}
\end{aligned}$$

($\boldsymbol{P}_Z$ ist symmetrisch und idempotent, d.h. $\boldsymbol{P}_Z = \boldsymbol{P}_Z'$ und $\boldsymbol{P}_Z\boldsymbol{P}_Z = \boldsymbol{P}_Z$)
Die geschätzte Varianz-Kovarianzmatrix des Koeffizientenvektors $\hat{\boldsymbol{\beta}}$ ist

$$\begin{aligned}
\widehat{\text{var}}(\hat{\boldsymbol{\beta}}) &= \hat{\sigma}^2(\widehat{\boldsymbol{X}}'\widehat{\boldsymbol{X}})^{-1} \\
&= \hat{\sigma}^2(\boldsymbol{X}'\boldsymbol{P}_Z'\boldsymbol{P}_Z\boldsymbol{X})^{-1} \\
&= \hat{\sigma}^2(\boldsymbol{X}'\boldsymbol{P}_Z\boldsymbol{X})^{-1}
\end{aligned}$$

Achtung 1: Die IV Residuen, die u.a. zur Berechnung von $\hat{\sigma}^2$ benötigt werden, werden als

$$\hat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^{\text{IV}}$$

berechnet, und nicht wie manchmal irrtümlich angenommen aus $\boldsymbol{y} - \boldsymbol{Z}\hat{\boldsymbol{\beta}}^{\text{IV}}$ oder $\boldsymbol{y} - \widehat{\boldsymbol{X}}\hat{\boldsymbol{\beta}}^{\text{IV}}$!

Achtung 2: Die meisten Programme geben für Instrumentvariablenschätzer ein Bestimmtheitsmaß R^2 aus, aber dieses R^2 darf *nicht* wie beim OLS-Schätzer als Anteil der durch die Regressoren erklärten Variation von y interpretiert werden! Außerdem kann das R^2 von Instrumentvariablenschätzungen negativ sein!

Instrumentvariablenschätzer und die Methode 2-stufigen Kleinsten Quadrate (2SLS)

Da $\boldsymbol{P}_Z$ eine Projektionsmatrix ist kann der vorhin entwickelte IV Schätzer auch als eine zweistufige Anwendung des OLS-Schätzers begriffen werden

- **1. Stufe:** Regressiere jede der Variablen in der X-Matrix auf die Instrumente $\boldsymbol{Z}$ und forme daraus die Matrix der gefitteten Werte $\widehat{\boldsymbol{X}}$

  $$\widehat{\boldsymbol{X}} = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X} = \boldsymbol{P}_Z\boldsymbol{X}$$

  mit $\boldsymbol{P}_Z = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'$, wobei die Projektionsmatrix $\boldsymbol{P}_Z$ symmetrisch und idempotent ist.

- **2. Stufe:** Regressiere $\boldsymbol{y}$ auf die gefitteten $\widehat{\boldsymbol{X}}$ um um einen Schätzer $\hat{\boldsymbol{\beta}}$ für $\boldsymbol{\beta}$ zu erhalten

  $$\begin{aligned}
  \hat{\boldsymbol{\beta}}_{\text{2SLS}} &= (\widehat{\boldsymbol{X}}'\widehat{\boldsymbol{X}})^{-1}(\widehat{\boldsymbol{X}}'\boldsymbol{y}) \\
  &= [\boldsymbol{X}'\boldsymbol{P}_Z'\boldsymbol{P}_Z\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{P}_Z'\boldsymbol{y} \\
  &= [\boldsymbol{X}'\boldsymbol{P}_Z\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{P}_Z\boldsymbol{y} \\
  &= \left[\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}\right]^{-1}\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{y} \\
  &= \hat{\boldsymbol{\beta}}^{\text{IV}}
  \end{aligned}$$

Deshalb ist der 2SLS Schätzer ein Spezialfall des IV Schätzers.

10.7 Tests auf Endogenität und überidentifizierende Restriktionen

Für die Anwendung der IV-Methode sind zwei Fragen wichtig

- ist eine erklärende Variable der Strukturgleichung tatsächlich endogen? Diese Frage kann mit Hilfe eines *Hausman Tests* getestet werden. Sollte dieser Test Hinweise auf Endogenität liefern kann z.B. ein IV-Schätzer verwendet werden.
- wenn auf einen IV-Schätzer zurückgegriffen werden muss, sind die verwendeten Instrumente tatsächlich exogen? Ein Test dieser Hypothese ist nur in überidentifizierten Gleichungen möglich (*Test auf überidentifizierende Restriktionen*, z.B. Sargan Test).

Hausman Test (Test auf Endogenität)

Da die Auswirkungen einer Korrelation zwischen den erklärenden Variablen und den Residuen der Grundgesamtheit sehr weitreichend sind (d.h. der Schätzer wäre systematisch verzerrt) wäre ein Test auf diese Korrelation sehr hilfreich. Allerdings können dafür nicht die aus der Stichprobe geschätzten Residuen $\hat{\varepsilon}_i$ herangezogen werden, da diese per Konstruktion immer mit den x unkorreliert sind (diese Orthogonalität ist eine Bedingung erster Ordnung!).

Hausman (1978) hat dafür einen sehr einfachen asymptotischen Test vorgeschlagen. Die entsprechende Null- und Alternativhypothese ist

$$\mathrm{H}_0: \ \operatorname{plim} \frac{1}{n} \sum (x_i - \bar{x})\varepsilon_i = 0 \qquad \mathrm{H}_1: \ \operatorname{plim} \frac{1}{n} \sum (x_i - \bar{x})\varepsilon_i \neq 0$$

Die Grundidee des Hausman Tests besteht in einem Vergleich zweier Schätzer, wovon einer sowohl unter der Null-Hypothese als auch unter der Alternativhypothese konsistent ist, während der zweite nur unter Nullhypothese konsistente Schätzresultate liefert. Ein großer Unterschied zwischen diesen beiden Schätzungen wird als Evidenz zugunsten der Alternativhypothese gedeutet.

Hausman hat gezeigt, dass unter Gültigkeit der Nullhypothese die folgende Statistik asymptotisch χ^2 verteilt ist mit einem Freiheitsgrad

$$H = \frac{\left(\widehat{\beta}_h^{\mathrm{IV}} - \widehat{\beta}_h^{\mathrm{OLS}}\right)^2}{\widehat{\operatorname{var}}\left(\widehat{\beta}_h^{\mathrm{IV}}\right) - \widehat{\operatorname{var}}\left(\widehat{\beta}_h^{\mathrm{OLS}}\right)} \overset{a}{\sim} \chi_1^2$$

wobei $\widehat{\beta}_h^{\mathrm{IV}}$ der Instrumentvariablen und $\widehat{\beta}_h^{\mathrm{OLS}}$ der OLS-Schätzer ist.
Für die Berechnung der Varianzen sollten in beiden Fällen die gleiche Schätzung für $\hat{\sigma}^2$ herangezogen werden!
Wir verwerfen die Nullhypothese und schließen auf eine kontemporäre Korrelation zwischen x_i und ε_i wenn $H > \chi_c^2$ (für ein 5% Signifikanzniveau ist der kritische Wert $\chi_c^2 = 3.84$).

Hausman-Tests werden vor allem verwendet für die Erkennung von

- Simulanitäts-Bias
- Messfehlern
- Omitted Variables

oder allgemeiner, in Fällen in denen die Störterm ε mit erklärenden Variablen x korreliert ist.

Beispiel (Stata): Für das Beispiel mit der Lernzeit erhalten wir

```
. insheet using "https://www.uibk.ac.at/econometrics/data/iv_bsp_lz.csv"
. ivregress 2sls pkt (lz = w)
. estimates store tsls
. regress pkt lz
. estimates store ols
. hausman tsls ols

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |      tsls         ols         Difference          S.E.
-------------+----------------------------------------------------------------
          lz |    7.350636    -1.700709        9.051345        2.608475
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from ivregress
          B = inconsistent under Ha, efficient under Ho; obtained from regress

    Test:  Ho:  difference in coefficients not systematic
                  chi2(1) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =       12.04
                Prob>chi2 =      0.0005
```

Wir würden die Nullhypothese also verwerfen und (korrekt) schließen, dass die Lernzeit ein endogener Regressor ist.

Eine einfache Version dieses Tests mittels Hilfsregression wurde von Davidson and MacKinnon (1989) vorgeschlagen. Die Idee beruht einfach darauf, dass eine z.B. möglicherweise fehlerhaft gemessene Variable in einer ersten Stufe auf ein Instrument regressiert wird, und die Residuen dieser Hilfsregression in der ursprünglichen Regression als zusätzlicher Regressor berücksichtigt wird. Wenn der Koeffizient dieser Residuen signifikant von Null verschieden ist wird dies als Hinweis dafür interpretiert, dass die eine Korrelation zwischen x und ε (möglicherweise hervorgerufen durch Messfehler, fehlende Variablen, Endogenität) einen Einfluss auf das Schätzergebnis hat. Da es ein asymptotischer Test ist wird die Normalverteilung angenommen (bzw. wenn mehrere Variablen getestet werden die χ^2 Verteilung), aber es hat sich eingebürgert die meist einfacher im Regressionsoutput verfügbare t-Statistik (bzw. F-Statistik) heranzuziehen. Aufgrund der asymptotischen Gültigkeit macht dies keinen großen Unterschied (für nähere Hinweise zum Wu-Hausman Test siehe z.B. Johnston, DiNardo 1997, S. 257ff).

Angenommen, das Modell in Strukturform sei

$$y_1 = \beta_1 + \beta_2 y_2 + \beta_4 x_1 + \beta_5 x_2 + \varepsilon$$

und wir vermuten, dass y_2 endogen ist, d.h. mit ε korreliert ist. Wenn wir für y_2 über ein Instrument z verfügen gehen wir 2-stufig vor

1. Man schätzt eine *reduzierte Form* für y_2

   $$y_2 = \hat{\alpha}_0 + \hat{\alpha}_1 x_1 + \hat{\alpha}_2 x_2 + \hat{\alpha}_3 z + \hat{\nu}$$

   und berechnet daraus die Residuen $\hat{\nu}$.

2. Man schätzt die ursprüngliche Strukturform und verwendet die Stichprobenresiduen der reduzierten Form als *zusätzlichen* Regressor.

   $$y_1 = \widehat{\beta}_1 + \widehat{\beta}_2 y_2 + \widehat{\beta}_4 x_1 + \widehat{\beta}_5 x_2 + \widehat{\beta}_6 \hat{\nu} + \tilde{\varepsilon}$$

   Wenn $\widehat{\beta}_6$ signifikant von Null verschieden ist schließen wir daraus, dass y_2 tatsächlich endogen ist, OLS also verzerrte Ergebnisse liefert.

Dieser Test kann einfach für mehrere endogene Variablen verallgemeinert werden. Man berechnet für jede endogene Variable die Residuen der reduzierten Form, verwendet diese als zusätzliche Regressoren in der Strukturform, und testet deren gemeinsame Signifikanz mit Hilfe eines einfachen F-Tests.

Test auf überidentifizierende Restriktionen

Instrumente müssen *relevant* und *exogen* sein. Während ein F-Test auf die gemeinsame Signifikanz aller Instrumente in der 1. Stufe Hinweise auf die Relevanz der Instrumente liefert kann die Korrelation zwischen den Instrumenten und den unbeobachtbaren Störtermen der Grundgesamtheit (d.h. die Exogenität) nicht unmittelbar getestet werden.

Allerdings gibt es einen Test für überidentifizierte Systeme. Angenommen, die Strukturgleichung sei

$$y_{i1} = \beta_1 + \beta_2 y_{i2} + \beta_3 x_i + \varepsilon_i$$

und wir verfügen über zwei potentielle Instrumentvariablen z_1 und z_2.

Wir könnten diese Gleichung mit nur einer Instrumentvariable 2-stufig schätzen, z.B. mit z_1, daraus die Residuen $\hat{\varepsilon}_1$ berechnen und überprüfen, ob diese mit der nicht verwendeten Instrumentvariable z_2 korreliert sind. Wenn $\mathrm{corr}(\hat{\varepsilon}_1, z_2) \neq 0$ wäre offensichtlich z_1 *oder* z_2 ein schlechtes Instrument! Umgekehrt könnte auch mit z_2 als Instrument begonnen werden und die Korrelation der Residuen dieser 2SLS Schätzung mit z_1 überprüft werden.

Ein einfacher Test auf überidentifizierende Restriktionen kann wieder mit Hilfe einer Hilfsregression durchgeführt werden:

1. Schätze die überidentifizierte Strukturgleichung mit 2SLS und berechne daraus die Stichproben-Residuen $\hat{\varepsilon}$, z.B.
$$y_{i1} = \widehat{\beta}_1 + \widehat{\beta}_2 \widehat{y}_{i2} + \widehat{\beta}_3 x_i + \hat{\varepsilon}_i \qquad \rightarrow \quad \text{Stichprobenresiduen } \hat{\varepsilon}$$
2. Schätze eine Hilfsregression von $\hat{\varepsilon}$ auf alle exogenen Variablen und Instrumente und bestimme daraus das Bestimmtheitsmaß R^2_*, z.B.
$$\hat{\varepsilon}_i = \hat{\alpha}_0 + \hat{\alpha}_1 z_{i1} + \hat{\alpha}_2 z_{i2} + \hat{\alpha}_3 x_i + \upsilon_i \qquad \rightarrow \qquad R^2_*$$
3. Unter der Nullhypothese, dass *alle* Instrumentvariablen exogen sind, ist
$$nR^2_* \overset{a}{\sim} \chi^2_q$$
wobei n die Anzahl der Beobachtungen ist. Die Anzahl der Freiheitsgrade q ist die Anzahl der Instrumentvariablen außerhalb des Modells abzüglich der Anzahl der endogenen erklärenden Variablen, d.h. der Grad der 'Überidentifikation'.

Wenn der Wert von nR^2_* größer ist als der entsprechende kritische Wert der χ^2-Verteilung muss die Nullhypothese, dass alle Instrumente exogen sind, verworfen werden.

Dieser Test ist ein Spezialfall des *Sargan-Tests* und ist nur für überidentifizierte Gleichungen anwendbar. Er gibt allerdings keine Hinweise darauf, *welche* Instrumente möglicherweise endogen sind!

Beispiele finden Sie im Abschnitt zu 'Simultaner Kausalität'.

10.8 Messfehler in den erklärenden (x) Variablen

Eine weitere mögliche Ursache für Endogenität sind Messfehler in mindestens einer x Variablen. Um dies zu zeigen nehmen wir an, dass anstelle des 'wahren' Wertes[6] $\ddot{x}_i$ nur ein fehlerhaft gemessenes

$$\ddot{x}_i^* = \ddot{x}_i + \upsilon_i$$

beobachtet werden kann ($\ddot{x}_i := x_i - \bar{x}$ bezeichne wieder Abweichungen vom Mittelwert), mit $\upsilon_i \sim \text{i.i.d.}(0, \sigma^2_{\upsilon_i})$.

Das wahre (und unbeobachtbare) Regressionsmodell sei

$$\begin{aligned} \ddot{y}_i &= \beta \ddot{x}_i + \varepsilon_i \\ &= \beta(\ddot{x}_i^* - \upsilon_i) + \varepsilon_i \\ &= \beta \ddot{x}_i^* + (\varepsilon_i - \beta \upsilon_i) \end{aligned}$$

aber wir können nur das Modell

$$\ddot{y}_i = \beta \ddot{x}_i^* + \varepsilon_i^* = \beta \underbrace{(\ddot{x}_i + \upsilon_i)}_{\ddot{x}_i^*} + \underbrace{(\varepsilon_i - \beta \upsilon_i)}_{\varepsilon_i^*}$$

[6]Einfachheitshalber beschränken wir uns im folgenden wieder auf den 2-Variablen Fall. Für eine allgemeinere Darstellung in Matrixschreibweise siehe z.B. Johnston 1997, S. 154ff.

schätzen. Schon hier erkennen wir, dass der Messfehler υ im Regressor $\ddot{x}^*$ und im Störterm ε^* auftaucht, weshalb diese korreliert sein werden, also Endogenität im ökonometrischen Sinne verursacht.

Wir beschränken uns auf den einfachsten Fall, wenn die Störterme alle angenehmen Eigenschaften aufweisen: $\mathrm{E}(\varepsilon_i) = \mathrm{E}(\upsilon_i) = 0$, $\mathrm{var}(\varepsilon_i) = \mathrm{E}(\varepsilon_i^2) = \sigma_\varepsilon^2$, $\mathrm{var}(\upsilon_i) = \mathrm{E}(\upsilon_i^2) = \sigma_\upsilon^2$, $\mathrm{cov}(\varepsilon_i, \varepsilon_j) = \mathrm{cov}(\upsilon_i, \upsilon_j) = 0$ für $i \neq j$, und $\mathrm{cov}(\varepsilon_i, \upsilon_j) = 0$ für alle i, j.

Selbst wenn die Störterme all diese angenehmen Eigenschaften aufweisen und mit den wahren Werten $\ddot{x}_i$ unkorreliert sind, so ist der OLS–Schätzer in der Regel dennoch *weder erwartungstreu noch konsistent!* Dies kann einfach gezeigt werden

Der OLS-Schätzer mit dem fehlerhaft gemessenen $\ddot{x}^*$ ist

$$\begin{aligned}
\widehat{\beta}_2 &= \frac{\sum \ddot{x}^* \ddot{y}}{\sum (\ddot{x}^*)^2} \\
&= \frac{\sum \ddot{x}^*(\beta \ddot{x} + \varepsilon)}{\sum (\ddot{x}^*)^2} \\
&= \beta \frac{\sum (\ddot{x}^*) \ddot{x}}{\sum (\ddot{x}^*)^2} + \frac{\sum \ddot{x}^* \varepsilon}{\sum (\ddot{x}^*)^2}
\end{aligned}$$

In diesem Fall können wir nicht einfach den Erwartungswert berechnen, da aufgrund des Messfehlers auch der Nenner des Schätzers stochastisch ist.[7] Aber man kann mit Hilfe des *probability limits* immerhin die *Konsistenz* überprüfen.

Wenn die entsprechenden Momente 2. Ordnung existieren und

$$\begin{aligned}
\mathrm{plim}\left(\frac{1}{n}\sum (\ddot{x}^*)^2\right) &= \sigma_{\ddot{x}}^2 + \sigma_\upsilon^2 \\
\mathrm{plim}\left(\frac{1}{n}\sum (\ddot{x}^*)\ddot{x}\right) &= \sigma_{\ddot{x}}^2 \\
\mathrm{plim}\left(\frac{1}{n}\sum \ddot{x}^* \varepsilon\right) &= 0
\end{aligned}$$

folgt daraus

$$\mathrm{plim}(\widehat{\beta}_2) = \mathrm{plim}\left(\beta \frac{\sum (\ddot{x}^*)\ddot{x}}{\sum (\ddot{x}^*)^2} + \frac{\sum \ddot{x}^* \varepsilon}{\sum (\ddot{x}^*)^2}\right) = \beta \left(\frac{\sigma_{\ddot{x}}^2}{\sigma_{\ddot{x}}^2 + \sigma_\upsilon^2}\right)$$

Da Varianzen immer positiv sind ist der Nenner von $\sigma_{\ddot{x}}^2/(\sigma_{\ddot{x}}^2 + \sigma_\upsilon^2)$ immer größer als der Zähler ($\sigma_\upsilon^2 > 0$). Deshalb führt der Messfehler im bivariaten Modell dazu, dass der OLS Steigungsparameter $\widehat{\beta}_2$ näher bei Null liegt als der wahre Wert β_2 (auf englisch wird dies *'attenuation bias'* genannt). Die *'attenuation'* gilt allerdings nur im bivariaten Modell, in multiplen Regressionsmodellen kann der Bias komplizierter sein.

Fehlerhaft gemessene Variable treten häufig auf, wenn Proxy-Variablen verwendet werden, da die eigentlich interessierenden Variable unbeobachtbar ist (z.B. das permanente Einkommen in einer Konsumfunktion nach Friedmann, oder die Begabung

[7] Erinnern Sie sich, $\mathrm{E}(X/Y) \neq \mathrm{E}(X)/\mathrm{E}(Y)$!

eines Arbeitnehmers), wenn Fragen in einem Fragebogen zweideutig sind, oder wenn Geräte ungenaue Daten liefern. In allen diesen Fällen ist mit systematisch verzerrten Schätzergebnissen zu rechnen!

Die folgende kleine Monte Carlo Simulation soll dies demonstrieren:

```
' Monte Carlo: errors in variables
!n = 300
!REP = 1000

wfcreate u !n
vector(!REP) R
rndseed 1234567
series x = @rnorm
for !r = 1 to !REP
  series u = @rnorm
  series y = 5 + 5*x + u   ' Datengen. Prozess
  series xs = x + @rnorm   ' x wird fehlerhaft gemessen
  equation eq.ls y c xs    ' xs = x + Fehler
  R(!r) = c(2)
next
R.distplot hist
```

Da in diesem einfachen Fall $\beta_2 = 5$, $x_i \sim \mathrm{N}(0,1)$, $\upsilon_i \sim \mathrm{N}(0,1)$ $\varepsilon_i \sim \mathrm{N}(0,1)$ ist

$$\operatorname{plim}(\widehat{\beta}_2) = \beta_2 \left(\frac{\sigma^2_{\ddot{x}}}{\sigma^2_{\ddot{x}} + \sigma^2_{\upsilon}} \right) = \beta_2 \, \frac{1}{1+1} = 0.5\beta_2 = 2.5$$

Der Mittelwert dieser tausend Replikationen (gespeichert in Vektor R) ist ungefähr 2.46 und die Standardabweichung 0.11, was denn Erwartungen entspricht. Abbildung 10.7 zeigt das Histogramm

Wenn geeignete Instrumentvariablen zur Verfügung stehen erlauben diese zumindest eine konsistente Schätzung.

Exkurs: Messfehler in der abhängigen Variablen *y* Die vorhergehenden Ausführungen bezogen sich ausschließlich auf Meßfehler in der erklärenden x Variable, Meßfehler in der abhängigen y Variable sind im Vergleich dazu verhältnismäßig unproblematisch. Um dies zu zeigen nehmen wir an, das wahre Regressionsmodell sei durch die Beziehung

$$\ddot{y}_i = \beta \ddot{x}_i + \varepsilon_i$$

gegeben, wobei ε_i ein üblicher Störterm sei. Wenn nun anstelle der 'wahren' abhängigen Variablen $\ddot{y}_i$ nur ein fehlerhaft gemessenes $\ddot{y}_i^*$ beobachtbar ist, wobei $\ddot{y}_i^* = \ddot{y}_i + \upsilon_i$ mit $\operatorname{cov}(\upsilon_i, \varepsilon_i) = 0$, dann berechnen wir die Regression

$$\ddot{y}_i^* = \beta \ddot{x}_i + (\varepsilon_i + \upsilon_i)$$

Eine OLS-Schätzung für β ist weiterhin unverzerrt und effizient. Die Varianz des Störterms $(\varepsilon_i + \upsilon_i)$ ist nun zwar größer, dem wird aber bei der Berechnung von $\hat{\sigma}^2$ aus

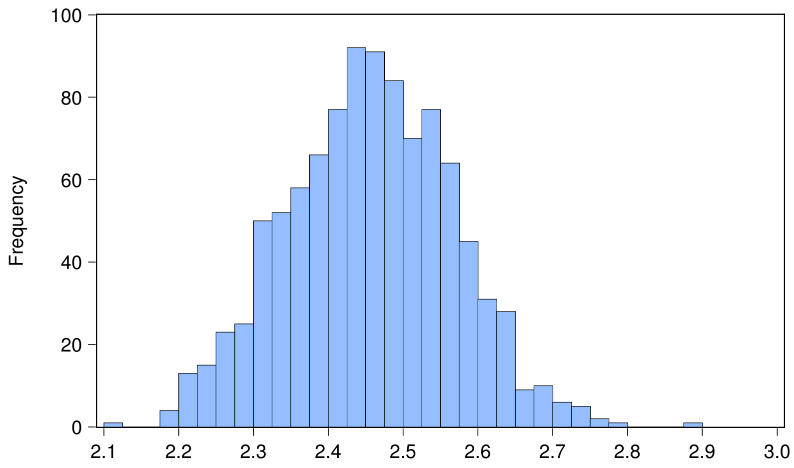


Abbildung 10.7: Monte Carlo Simulation für Messfehler im Regressor. Der wahre Wert ist $\beta_2 = 5$, aufgrund der Messfehler wird dieser Wert massiv unterschätzt.

den beobachteten Störtermen Rechnung getragen. Deshalb bleiben alle statistischen Tests gültig.

Systematische Verzerrungen treten nur auf, wenn die erklärenden x Variablen fehlerhaft gemessen werden.

10.9 Simultane Kausalität

Stellen Sie sich ein Haus mit Zentralheizung vor. Die Temperatur wird durch eine Ölheizung mit Thermostat geregelt, der die Innentemperatur (P) unabhängig von der Außentemperatur (V) konstant hält. Wir beobachten Innen- und Außentemperatur sowie den Ölverbrauch M. Wenn der Thermostat perfekt funktioniert beobachten wir, dass die Innentemperatur weder mit der Außentemperatur noch mit dem Ölverbrauch korreliert ist, denn die Innentemperatur bleibt konstant. Ein außerirdischer Ökonometriker, der die Rolle des Thermostats nicht kennt, würde folgerichtig schließen, dass die Zunahme des Ölverbrauchs zu einem Absinken der Außentemperatur führt, dass aber weder die Außentemperatur noch der Ölverbrauch in irgendeinem Zusammenhang mit der Innentemperatur stehen. Der einzige Effekt des Verbrennens von Öl scheint also in einem Absenken der Außentemperatur zu bestehen. Eine Kollegin des außerirdischen Ökonometrikes würde sich vielleicht fragen, ob die Kausalität nicht umgekehrt läuft, das heißt, ob nicht eine Zunahme der Außentemperatur zu einem Rückgang des Ölverbrauchs führt, aber sie würde ihrem Kollegen zustimmen, dass kein Zusammenhang zwischen dem Ölverbrauch und der Innentemperatur besteht. Sie würden dem Hausherren also einvernehmlich raten Heizkosten zu sparen und die Heizung abzustellen.

Natürlich hilft hier auch eine multiple Regression: $P = \widehat{\beta}_1 + \widehat{\beta}_2 V + \widehat{\beta}_3 M + \hat{\varepsilon}$ nicht weiter, die Außentemperatur V und der Ölverbrauch M sind vermutlich fast perfekt korreliert (multikollinear), aber leisten auch gemeinsam keinen Erklärungsbeitrag für die Innentemperatur P. Mit Beobachtungsdaten alleine ist es offensichtlich nicht möglich die Kausalitätsstruktur zwischen diesen Variablen zu erkennen.

Dieses Beispiel stammt von Milton Friedman (1912 – 2006) und sollte den Zusammenhang zwischen beobachteter Inflationsrate P , Geldmenge M, und Umlaufgeschwindigkeit des Geldes V erklären; die Rolle des Thermostats spielt natürlich die Zentralbank.

Dieses Beispiel hat allerdings eine viel breitere Bedeutung, wann immer Personen ihre Handlungen an die äußeren Bedingungen anpassen haben wir ein sehr ähnliches Problem, sie reagieren wie der Thermostat.

Das grundlegende Problem in diesem Beispiel ist, dass wir für den Thermostat keine Gleichung haben, und deshalb heben wir den datengenerierenden Prozess mit einer einzelnen Gleichung falsch spezifiziert.

Ähnlich wie wir früher von einem *'omitted variables bias'* gesprochen haben könnten wir hier von einem *'omitted equation bias'* sprechen.

Was passiert aber, wenn die Variablen auf komplexere Weise verknüpft sind, so dass wir zur Beschreibung mindestens *zwei oder mehrere Gleichungen* benötigen? Dies ist in den Sozialwissenschaften eher die Regel als die Ausnahme, selbst zur Beschreibung des einfachsten Marktes benötigen wir bereits eine Nachfrage- *und* Angebotsfunktion!

Wir werden nun zeigen, dass eine solche gegenseitige Abhängigkeit, zu deren Beschreibung mehr als eine Gleichung benötigt wird, ebenfalls zu einer Korrelation zwischen dem Störterm und den erklärenden x Variablen führt.

Der einfachste Fall einer solchen Korrelation zwischen Regressor und Störterm kann einfach anhand des keynesianischen Eingaben-Ausgaben Modells mit einer stochastischen Konsumfunktion und Einkommensidentität erläutert werden.

Das der Abbildung 10.8 zugrunde liegende datengenerierende Prozess ist

$$\begin{aligned} C_i &= 60 + 0.5Y_i + \varepsilon_i \qquad \text{mit } \varepsilon_i \sim \mathrm{U}(-30, +30) \\ Y_i &= C_i + I_i \qquad \text{mit } I_i \sim \mathrm{U}(30, 80) \end{aligned}$$

wobei U hier für die Gleichverteilung steht, $I_i \sim \mathrm{U}(30, 80)$ bedeutet also, dass die Variable I_i mit gleicher Wahrscheinlichkeit irgendeinen Wert zwischen 30 und 80 annimmt. Die Gleichverteilung wurde nur gewählt um eine übersichtlichere Grafik zu erhalten und spielt ansonsten keine Rolle.

Abbildung 10.8 zeigt diese Funktionen für die Extremwerte von ε_i und I_i (d.h. $C_i = 60 + 0.5Y_i - 30$ und $C_i = 60 + 0.5Y_i + 30$, bzw. $C_i = Y_i - 30$ und $C_i = Y_i - 80$; die strichlierten Linen zeigen die Funktionen für $\varepsilon_i = 0$ und $I_i = 0$).

Die Daten wurden diesem Modell entsprechend vom Computer erzeugt. Eine einfache OLS Schätzung der Konsumfunktion liefert statt der wahren Werte $\widehat{\beta}_1 = 60$ und $\widehat{\beta}_2 = 0.5$ völlig andere Werte.

$$\begin{aligned} \mathrm{C} \;=\; &\underset{(5.685)}{-9.175} \;+\; \underset{(0.024)^{***}}{0.808}\ \mathrm{Y} \end{aligned}$$

$$R^2 = 0.934, \quad s = 10.818, \quad F\text{-Stat} = 1096.971, \quad n = 80$$

(Standardfehler in Klammern)

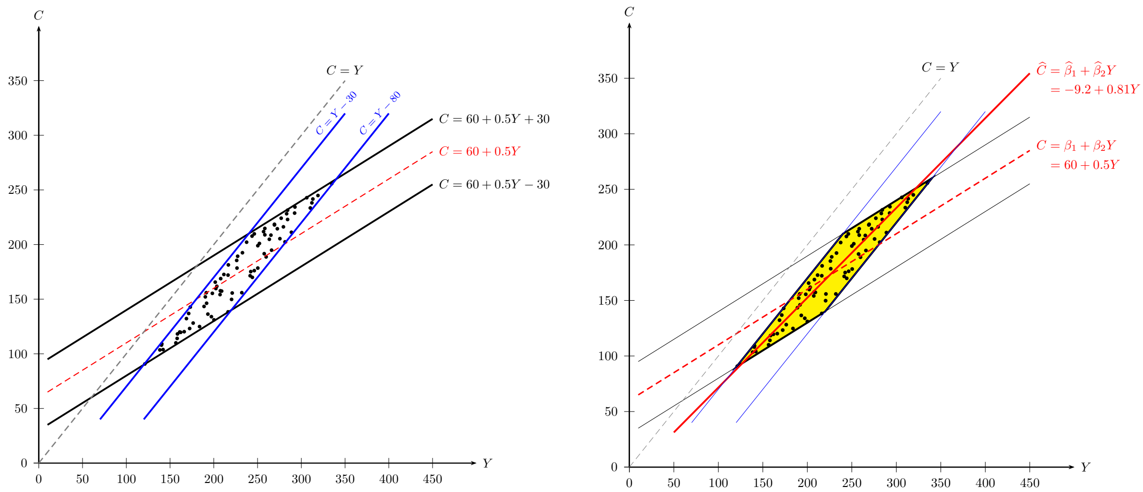


Abbildung 10.8: Ein einfaches Keynesianisches Modell.

Abbildung 10.8 zeigt, was schief gelaufen ist, die Einkommensidentität $Y = C + I$ erzwingt, dass alle Realisationen im gelben Parallelogramm liegen müssen!

Wie Abbildung 10.9 zeigt führt dies zu einer Korrelation zwischen den Störtermen und der erklärenden Variable Y, denn ein in einer Periode zufällig größerer Störterm führt zu erhöhten Konsumausgaben, diese über die Identität zu einem höheren Einkommen, weshalb Einkommen und Störterme korreliert sind.

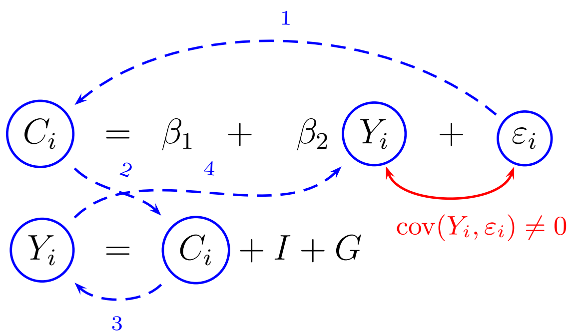


Abbildung 10.9: Simultanität im keynesianischen Eingaben-Ausgaben Modell [local, www]

Wie das rechte Panel von Abbildung 10.8 zeigt, liefert die OLS Gerade zwar eine optimale Beschreibung der realisierten Daten, aber sie liefert keine konsistente Schätzung der Konsumfunktion, weil die Abhängigkeit zwischen Konsum und Einkommen *über die Einkommensidentität* nicht berücksichtigt wird!

Die OLS Koeffizienten $\widehat{\beta}_1$ und $\widehat{\beta}_2$ liefern deshalb keine unverzerrten Schätzer für den autonomen Konsum β_1 und die marginale Konsumneigung β_2! Da dieses Problem unabhängig von der Stichprobengröße existiert wird das Problem mit zunehmender Stichprobengröße auch nicht kleiner, deshalb ist der OLS Schätzer in diesem Fall auch nicht konsistent!

Eine einfache Monte Carlo Simulation soll das Problem wieder veranschaulichen. Das folgende kleine EViews Programm erzeugt tausend Mal ein Modell mit zwei

Gleichungen, einer Konsumfunktion und einer Identität

$$\begin{aligned} \mathrm{C}_i &= 60 + 0.5Y + \varepsilon_i \\ Y_i &= \mathrm{C}_i + Z_i \end{aligned}$$

mit $\varepsilon_i \sim \mathrm{U}(-30, +30)$ und $Z_i \sim \mathrm{U}(30, 80)$, wobei Z alle exogenen Nachfragekategorien enthält (zum Beispiel Investitionen und Staatsausgaben).

Das folgende EViews-Programm löst dieses Modell bei jedem der tausend Durchgänge neu und schätzt jedes Mal *aus der Lösungen* die marginale Konsumneigung. Man beachte, dass die 'wahre' marginale Konsumneigung mit 0.5 vorgegeben wurde.

```
wfcreate u 25
rndseed 123456789
!REP = 1000
vector(!REP) MCP = na
series Cons
series Y
series Z
for !r = 1 to !REP
Z = @runif(30,80)
    ' Modell lösen
    model Keynes
    Keynes.append  Cons = 60 + 0.5*Y + @runif(-30,30)
    Keynes.append  Y = Cons + Z
    Keynes.solve
    delete Keynes
    ' Konsumfunktion mit Modelllösungen schätzen
    equation temp.ls Cons_0 c Y_0
    MCP(!r) = c(2)
next
MCP.distplot hist
```

Abbildung 10.10 zeigt das Histogramm der tausend Schätzungen, offensichtlich ist die Schätzung der marginalen Konsumneigung systematisch verzerrt, obwohl der wahre Wert der marginalen Konsumneigung mit 0.5 vorgegeben wurde, liegt der Mittelwert der empirischen Stichprobenkennwertverteilung des OLS Schätzers bei 0.8 (tausend Replikationen).

In diesem Fall gibt die Modellstruktur eine ideale Instrumentvariable vor, das exogene Z kommt in der Konsumgleichung nicht vor, aber Z ist über die Identität mit dem endogenen Regressor Y korreliert (ist also relevant). Abbildung 10.11 zeigt wieder das Histogramm der 1000 IV Schätzungen, offensichtlich liegt der Mittelwert deutlich näher bei dem wahren 0.5 (obwohl die Streuung deutlich größer ist).

Wann liegt Interdependenz vor? Häufig herrscht Verwirrung, wann tatsächlich von Interdependenz oder *'reverse causality'* gesprochen werden kann.

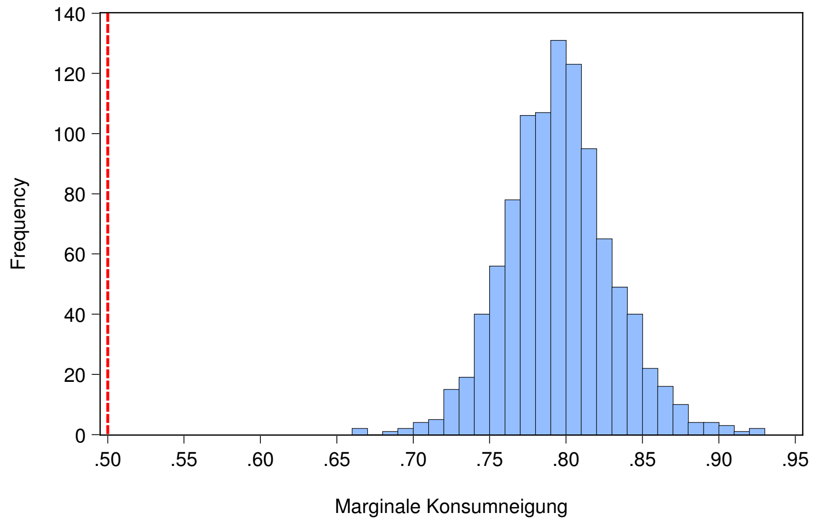


Abbildung 10.10: Monte Carlo Simulation für endogenen Regressor. Der wahre Wert der marginalen Konsumneigung ist 0.5; aufgrund des *feedbacks* über die Einkommensidentität überschätzt OLS diesen Wert systematisch.

Nehmen wir zum Beispiel an, jemand möchte die Konsumausgaben C eines Haushaltes untersuchen, und erklärt diese u.a. mit den Ersparnissen (S) und dem Einkommen (Y) des Haushaltes, also

$$C = \beta_1 + \beta_2 S \ + \beta_3 Y + \ldots \text{ andere Faktoren}$$

Eine andere Forscherin interessiert sich für die Determinanten der Ersparnisbildung, die u.a. von den Konsumausgaben und vom Einkommen abhängt, also

$$S = \alpha_1 + \alpha_2 C \ + \alpha_3 Y \ldots \text{ andere Faktoren}$$

Handelt es sich dabei um ein simultanes System? Nein! Die beiden Gleichungen modellieren das Entscheidungsverhalten des *gleichen* Haushaltes, beide Gleichungen enthalten als die selben Variablen, sie haben keine unabhängige Interpretation! Die zweite Gleichung ist lediglich eine Umformung der ersten Gleichung.

Angenommen jemand möchte versuchen die Auswirkungen der Anzahl von Polizisten (NP) in einer Region auf die Zahl der Diebstähle (D) zu untersuchen. Dazu schätzt er eine Regression

$$D = \beta_1 + \beta_2 \text{NP} \ + \ldots \text{ andere Faktoren}$$

Aber natürlich werden in Regionen mit höherer Kriminalität mehr Polizisten stationiert werden, weshalb

$$\text{NP} = \alpha_1 + \alpha_2 D \ + \ldots \text{ andere Faktoren}$$

Handelt es sich hierbei um ein simultanes System? Ja! Die erste Gleichung beschreibt das Verhalten von Dieben, während die zweite Gleichung das Verhalten von

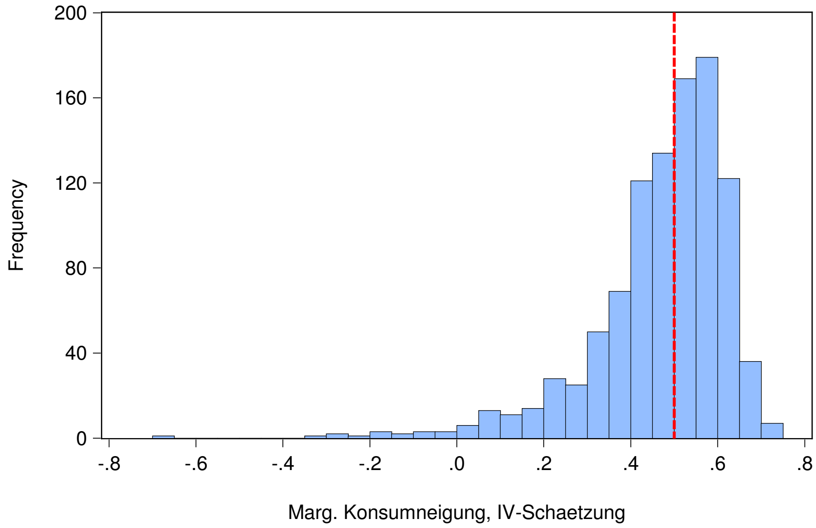


Abbildung 10.11: Monte Carlo Simulation für endogenen Regressor, IV-Schätzung mit Instrument Z. Der wahre Wert der marginalen Konsumneigung ist 0.5; der Mittelwert der tausend Replikationen ist 0.469.

Politikern oder Behörden beschreibt. Deshalb ist dies ein interdependentes Strukturmodell, das mit üblichen Methoden untersucht werden kann.

Im Kern tritt diese Problem immer dann auf, wenn zwei Variablen auf *mehr* als eine Art verknüpft sind, das heißt, wenn zur Beschreibung des Verhaltens einer Variable mehr als eine Gleichung benötigt wird. Das klassische Beispiel sind natürlich Angebots- und Nachfragefunktionen oder Makromodelle, aber ähnliche Probleme existieren fast überall in den Sozialwissenschaften.

Strukturform und reduzierte Form: Eine *Strukturgleichung* bildet einen theoretisch begründeten Zusammenhang ab. In einem simulatanen Strukturmodell (*'linear simultaneous equation model'*, SEM) sollte jede Gleichung im System eine eigenständige kausale Interpretation haben.[8]

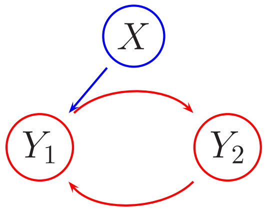
Durch Lösung eines Modells in Strukturform nach den endogenen Variablen erhält man die ***reduzierte Form*** des Gleichungssystems, in der es per Definition keine 'feed backs' mehr gibt.

In einer *komparativ-statischen Analyse* wird die reduzierte Form nach einer exogenen Variable abgeleitet. Da es in einer reduzierten Form keine 'feed backs' mehr gibt, sagt uns diese Ableitung, wie sich die Änderung der exogenen Variable *ceteris paribus*

[8]Sargan (1988, 27) definiert ein Modell, bzw. eine Struktur im Sinne der Cowles Commission etwas allgemeiner "A model is the specification of the probability distribution for a set of observations. A structure is the specification of the parameters of that distribution. Therefore, a structure is a model in which all the parameters are assigned numerical values." (zitiert nach Cameron and Trivedi, 2005, 20)

Strukturform:

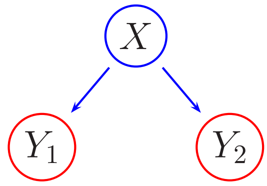
$$\begin{aligned} Y_1 &= Y_1(Y_2, X) \\ Y_2 &= Y_2(Y_1) \end{aligned}$$



⇒ Interdependenz

Reduzierte Form:

$$\begin{aligned} Y_1^* &= Y_1(X) \\ Y_2^* &= Y_2(X) \end{aligned}$$



⇒ keine feed-backs!

Abbildung 10.12: Strukturform und reduzierte Form eines Gleichungssystems

auf die gleichgewichtige endogene Variable auswirkt (in der Makroökonomik wird eine solche Analyse häufig *Multiplikatoranalyse* genannt).

Allerdings haben die Koeffizienten der reduzierten Form nur selten eine unmittelbare ökonomische Interpretation.

10.9.1 Beispiel: Fulton Fischmarkt

Für eine empirische Marktanalyse erhob Graddy (2006)[9] am Fulton Fischmarkt in New York vom 2.12.1991 – 8.5.1992 (111 Tage) täglich die Durchschnittspreise und Mengen für die Fischsorte Wittling (*'whiting'*, auch Merlan, eine Dorsch-Art), siehe auch Graddy (1995).

In einem ersten Ansatz könnte man eine log-lineare Regressionsfunktion mit konstanten Koeffizienten schätzen (mit $Q = \ln(\text{Quantity})$ und $P = \ln(\text{Price})$)

$$Q_t^{\text{obs}} = \beta_1^o + \beta_2^o P_t^{\text{obs}} + \varepsilon_t$$

$$\begin{array}{rcllcl} Q^{\text{obs}} & = & 8.419 & & - & 0.541\ P^{\text{obs}} \\ & & (0.076)^{***} & & & (0.179)^{***} \end{array}$$

$R^2 = 0.078, \quad s = 0.716, \quad F\text{-Stat} = 9.167, \quad \text{DW} = 1.556, \quad n = 111$
(Standardfehler in Klammern)

Ist dies eine Nachfragefunktion? Kann diese Funktion verwendet werden um Auswirkungen einer Steuer zu analysieren? Offensichtlich nicht, es ist nicht klar, ob es sich dabei um einen Nachfrage- oder Angebotsfunktion handelt.

Bei einer Nachfragefunktion

$$Q_t^d(P) = \beta_1^d + \beta_2^d P + \varepsilon_t^d$$

wird jedem P wird ein Q^d zugeordnet, dieser Zusammenhang folgt aus Präferenzen und Restriktionen; aber diese Nachfragefunktion ist *nicht beobachtbar*!

[9] `https://www.aeaweb.org/atypon.php?doi=10.1257/jep.20.2.207`

Diese Nachfragefunktion ist keine Regressionsfunktion, sondern eine *Strukturgleichung*. Sie beschreibt einen unbeobachtbaren Zusammenhang, in der Sprache der Treatment-Literatur ‘*‘potential outcomes’*.

Man beachte, dass P in der Nachfragefunktion keinen Subindex t aufweist, dies impliziert die starke Annahme, dass die Auswirkungen einer Änderung von P um eine Einheit auf Q_t^d für alle Preise und Markttage gleich β_2^d sind.

Die unbeobachtete Komponente ε_t^d ist beobachtungsspezifisch und sollte im Erwartungswert $\mathrm{E}(\varepsilon_t^d) = 0$.

Ähnliches gilt für die Angebotsfunktion

$$Q_t^s(P) = \beta_1^s + \beta_2^s P + \varepsilon_t^s$$

mit $\mathrm{E}[Q_t^s(P)] = \beta_1^s + \beta_2^s P$

Marktgleichgewicht

Sowohl Nachfrage- als auch Angebotsfunktion sind unbeobachtbar, aber wenn der Marktmechanismus funktioniert beobachten wir für jeden Tag/Markt ein $(P_t^{\text{obs}}, Q_t^{\text{obs}})$ Paar.

Im Gleichgewicht gilt

$$Q_t^{\text{obs}} = Q_t^d(P_t^{\text{obs}}) = Q_t^s(P_t^{\text{obs}})$$

Falls ein Gleichgewicht existiert und eindeutig ist können wir die (unbeobachtbaren) Angebots- und Nachfragefunktionen lösen und die beobachteten Mengen und Preise als Funktion der unbeobachteten Parameter und Störterme darstellen.

$$\begin{aligned} P_t^{\text{obs}} &= \frac{\beta_1^d - \beta_1^s}{\beta_2^s - \beta_2^d} + \frac{\varepsilon_t^d - \varepsilon_t^s}{\beta_2^s - \beta_2^d} \\ Q_t^{\text{obs}} &= \frac{\beta_2^s \beta_1^d - \beta_2^d \beta_1^s}{\beta_2^s - \beta_2^d} + \frac{\beta_2^s \varepsilon_t^d - \beta_2^d \varepsilon_t^s}{\beta_2^s - \beta_2^d} \end{aligned}$$

Dieses Modell sollte erklären, wie die beobachteten Mengen und Preise von den Marktbedingungen beeinflusst werden. Allerdings wird damit nicht der Mechanismus erklärt, wie (und ob) Marktpreise zustande kommen (z.B. Auktionator), das Gleichgewicht wird einfach angenommen.

Außerdem werden wir gleich sehen, dass diese Funktionen nicht geschätzt werden können.

Aber vorher ist es nützlich zu überlegen, was wir mit der ursprünglichen Regression

$$Q_t^{\text{obs}} = \beta_1^o + \beta_2^o P_t^{\text{obs}} + \varepsilon_t$$

eigentlich geschätzt haben.

Der Steigungskoeffizient dieser Regression ist

$$\beta_2^o = \frac{\operatorname{cov}(Q_t^{\text{obs}}, P_t^{\text{obs}})}{\operatorname{var}(P_t^{\text{obs}})}$$

Die Kovarianz zwischen beobachteten Gleichgewichtsmengen und -preisen ist $\mathrm{cov}(Q_t^{\mathrm{obs}}, P_t^{\mathrm{obs}}) = (\beta_2^s \sigma_d^2 + \beta_2^d \sigma_s^2 - 2\rho\sigma_d\sigma_s(\beta_2^d + \beta_2^s))/((\beta_2^s - \beta_2^d)^2)$, wobei σ_d^2 und σ_s^2 die Varianz von ε_t^d sowie ε_t^s sind und ρ die Korrelation zwischen ε_t^d und ε_t^s ist.

Die Varianz von P_t^{obs} ist $\mathrm{var}(P_t^{\mathrm{obs}}) = (\sigma_s^2 + \sigma_d^2 - 2\rho\sigma_d\sigma_s)/(\beta_2^s - \beta_2^d)^2$, siehe Imbens (2014, 19).

Wenn wir $\rho = 0$ annehmen erhalten wir deshalb als Steigungskoeffizient der Regression

$$\beta_2^o = \frac{\mathrm{cov}(Q_t^{\mathrm{obs}}, P_t^{\mathrm{obs}})}{\mathrm{var}(P_t^{\mathrm{obs}})} = \beta_2^s \frac{\sigma_d^2}{\sigma_s^2 + \sigma_d^2} + \beta_2^d \frac{\sigma_s^2}{\sigma_s^2 + \sigma_d^2}$$

Offensichtlich ist die beobachtete Steigung β_2^o ein gewichtetes Mittel der Steigungen der Angebots- und Nachfargefunktion und die Gewichte hängen nur vom Verhältnis der Varianzen der Störterme von Angebots- und Nachfargefunktion ab. Wenn σ_d^2 klein ist im Verhältnis zu σ_s^2 liegt die Steigung näher bei der Steigung der Nachfragefunktion, andernfalls näher bei der Steigung der Angebotsfunktion.

Eine einfache Regression von beobachteten Preisen auf die beobachteten Mengen ist also nicht geeignet, um z.B. die Auswirkungen einer Steuer abzuschätzen.

Dazu wird ein *Strukturmodell* benötigt, wie z.B. obige Nachfrage- und Angebotsfunktion. Die zentrale Annahme dabei ist, dass sowohl Nachfrager als auch Anbieter nur auf den wahrgenommenen Preis (d.h. inkl. Steuer) reagieren, sie sind *invariant* in dem Sinne, dass die Koeffizienten des Strukturmodells durch die Einführung einer Steuer nicht beeinflusst werden.

Wenn bei einem Preis von einem Euro z.B. eine Steuer von 10 Prozent pro kg ($r = 0.1$) eingeführt und die Konsumentin deshalb 1.1 €bezahlen muss, dann nehmen wir an, dass es für die Konsumentin keine Rolle spielt, ob die 10 Cent an das Finanzministerium oder an den Verkäufer gehen. Wenn $P_t(r)$ der Preis ist, den der Verkäufer erhält, dann sollte dieser Preis die Gleichgewichtsbedingung

$$Q_t^d(P_t(r)(1+r)) = Q_t^s(P_t(r))$$

lösen. Für eine lineare Angebots- und Nachfragefunktion erhält man z.B. den Gleichgewichtspreis, den der Verkäufer erhält

$$P_t(r) = \frac{\beta_1^d - \beta_1^s}{\beta_2^s - (1+r)\beta_2^d} + \frac{\varepsilon_t^d - \varepsilon_t^s}{\beta_2^s - (1+r)\beta_2^d}$$

Analog kann die Auswirkung der Steuer auf die Gleichgewichtsmenge ermittelt werden

Wenn wir die Strukturparameter β_1^s, β_2^s, β_1^d und β_2^d kennen würden, könnten wir die Auswirkungen einer Steuer beurteilen, die noch gar nicht eingeführt wurde, für die also auch noch keine Beobachtungen existieren.

Eine einfache Regression P_t^{obs} auf Q_t^{obs} erlaubt keine solche Vorhersage, deren Steigung ist bloß ein gewichtetes Mittel der Steigungen von Angebots- und Nachfragefunktion.

Das Angebots- und Nachfragemodell ist ein *Strukturmodell* in dem Sinne, dass beide Funktionen das Verhalten unterschiedlicher Personen(gruppen) beschreiben, und dass die Gleichungen durch Interventionen, wie z.B. der Einführung einer Steuer, nicht beeinflusst werden (sehr wohl aber die Gleichgewichtswerte).

Die nächste Frage ist, ob die Strukturparameter β_1^s, β_2^s, β_1^d und β_2^d aus den beobachteten Daten schätzen können. Im bisherigen Modell lautet die Antwort Nein, da wir nur Gleichgewichtswerte beobachteten. Um die vier unbekannten Parameter schätzen zu können benötigen wir zusätzliche Information, und dies führt uns zum *Identifikationsproblem.*

Stata:

```
* Fulton Fishmarket Whiting
clear all
use "https://www.uibk.ac.at/econometrics/data/fulton.dta"

* DEMAND
* OLS (verzerrt!)
regress qty price day1 day2 day3 day4
estimates store ols

ivregress 2sls qty day1 day2 day3 day4 ///
   (price = windspd windspd2 cold rainy mixed stormy)
estimates store tsls

hausman tsls ols    /// ist price endogen?
estat overid        /// sind Instrumente gültig?


* SUPPLY
ivregress 2sls price windspd windspd2 cold rainy mixed stormy ///
    (qty = day1 day2 day3 day4)
estat overid
```

R:

```
# Fulton Fishmarket Whiting
library("foreign")
fish <- read.dta("https://www.uibk.ac.at/econometrics/data/fulton.dta")
library("AER")

demand <- ivreg(qty ~ price + day1 + day2 + day3 + day4 | windspd +
    windspd2 + cold + rainy + mixed + stormy + day1 + day2 + day3 + day4, data = fish)
summary(demand)

supply <- ivreg(price ~ qty + windspd + windspd2 + cold + rainy +
    mixed + stormy | windspd + windspd2 + cold + rainy + mixed +
    stormy + day1 + day2 + day3 + day4, data = fish)
summary(supply)

library(stargazer)
stargazer(demand,supply)
```

Tabelle 10.2

	Dependent variable:	
	qty	price
	(1)	(2)
price	-0.934^{***}	
	(0.345)	
day1	-0.012	
	(0.208)	
day2	-0.526^{**}	
	(0.202)	
day3	-0.563^{***}	
	(0.207)	
day4	0.100	
	(0.203)	
qty		0.051
		(0.125)
windspd		1.447
		(4.067)
windspd2		-0.229
		(0.705)
cold		0.047
		(0.078)
rainy		-0.004
		(0.093)
mixed		0.212^{**}
		(0.096)
stormy		0.389^{***}
		(0.143)
Constant	8.540^{***}	-3.080
	(0.156)	(5.872)
Observations	111	111
R^2	0.184	0.194
Adjusted R^2	0.145	0.139
Residual Std. Error	0.686 (df = 105)	0.354 (df = 103)
Note:		*p<0.1; **p<0.05; ***p<0.01

10.9.2 Das Identifikationsproblem

Wir haben das Identifikationsproblem schon früher erwähnt, zum Beispiel im Zusammenhang mit Instrumentvariablen (IV). Dort stellten wir fest, dass für eine IV Schätzung zumindest so viele Instrumente benötigt werden, wie viele potentiell endogene Variablen als erklärende Variablen in der Gleichung vorkommen, anderenfalls kann der IV Schätzer nicht berechnet werden. Da wir dort ausschließlich Einzelgleichungsschätzungen hatten, war dies relativ einfach überprüfbar; wenn in einem Einzelgleichungsmodell weniger Instrumentvariablen als endogene Regressoren zur Verfügung stehen führt dies zu perfekter Multikollinearität, und der Schätzer ist nicht berechenbar. Bei Mehrgleichungsmodellen ist die Überprüfung der Identifizierbarkeit leider etwas komplizierter.

Das Problem der Identifikation stellt sich schon vor der Schätzung, Identifizierbarkeit ist eine logische Vorbedingung für die Schätzung von Strukturmodellen, oder in anderen Worten, Identifikation ist ein mathematisches Problem, kein Schätzproblem!

Die Intuition von Identifikation in Mehrgleichungsmodellen kann mit mit Hilfe eines einfachen Nachfrage- Angebotsmodells relativ einfach dargestellt werden.

$$\begin{aligned} \text{Angebot:} \qquad Q_t &= \alpha_1 + \alpha_2 P_t + u_t \\ \text{Nachfrage:} \qquad Q_t &= \beta_1 + \beta_2 P_t + v_t \end{aligned}$$

mit $\alpha_2 > 0$, $\beta_2 < 0$ und $\alpha_1 < \beta_1$.

Wenn der Markt im Gleichgewicht ist können wir nur Schnittpunkte dieser beiden Geraden beobachten, und es gibt keine Möglichkeit die Parameter $\alpha_1, \alpha_2, \beta_1$ und β_2 zu schätzen, die Daten enthalten nicht genügend Information um die Parameter der Strukturform eindeutig bestimmen zu können.

Dies kann man auch einfach anhand von Abbildung 10.13 (Seite 50) erkennen, Panele a), b), c) und d) zeigen, dass jede Beobachtung der Schnittpunkt von einer Angebots- und Nachfragefunktion ist, und die Daten deshalb keine Information über den Anstieg von Angebots- bzw. Nachfragefunktion enthalten. Deshalb sind weder Angebots- noch Nachfragefunktion identifizierbar. In anderen Worten, die Daten sind mit beliebigen Parametern $\alpha_1, \alpha_2, \beta_1$ und β_2 kompatibel,

Die reduzierten Gleichungen für das System in 'Abweichungsform' (d.h. die Variablen sind als Abweichungen vom Mittelwert gemessen) sind:

$$\ddot{P}_t = \frac{v_t - u_t}{\alpha_2 - \beta_2} \qquad \text{und} \qquad \ddot{Q}_t = \frac{\alpha_2 v_t - \beta_2 u_t}{\alpha_2 - \beta_2}$$

Offensichtlich können beliebig viele Geraden durch die Schnittpunkte (Gleichgewichtpunkte) gelegt werden, d.h. es gibt unendlich viele strukturelle Modelle, die mit dem gleichen reduzierten System vereinbar sind! Dieses Problem ändert sich nicht wenn mehr Beobachtungen verfügbar werden, solange dies nur mehr vom Gleichen bedeutet.

Wenn eine nicht identifizierbarer Strukturgleichung geschätzt wird ist das Ergebnis nicht kausal interpretierbar, trotzdem dürfte dies in der Vergangenheit häufiger vorkommen sein, auch, aber nicht nur, im Marketingbereich, wie das folgende Zitat belegt:

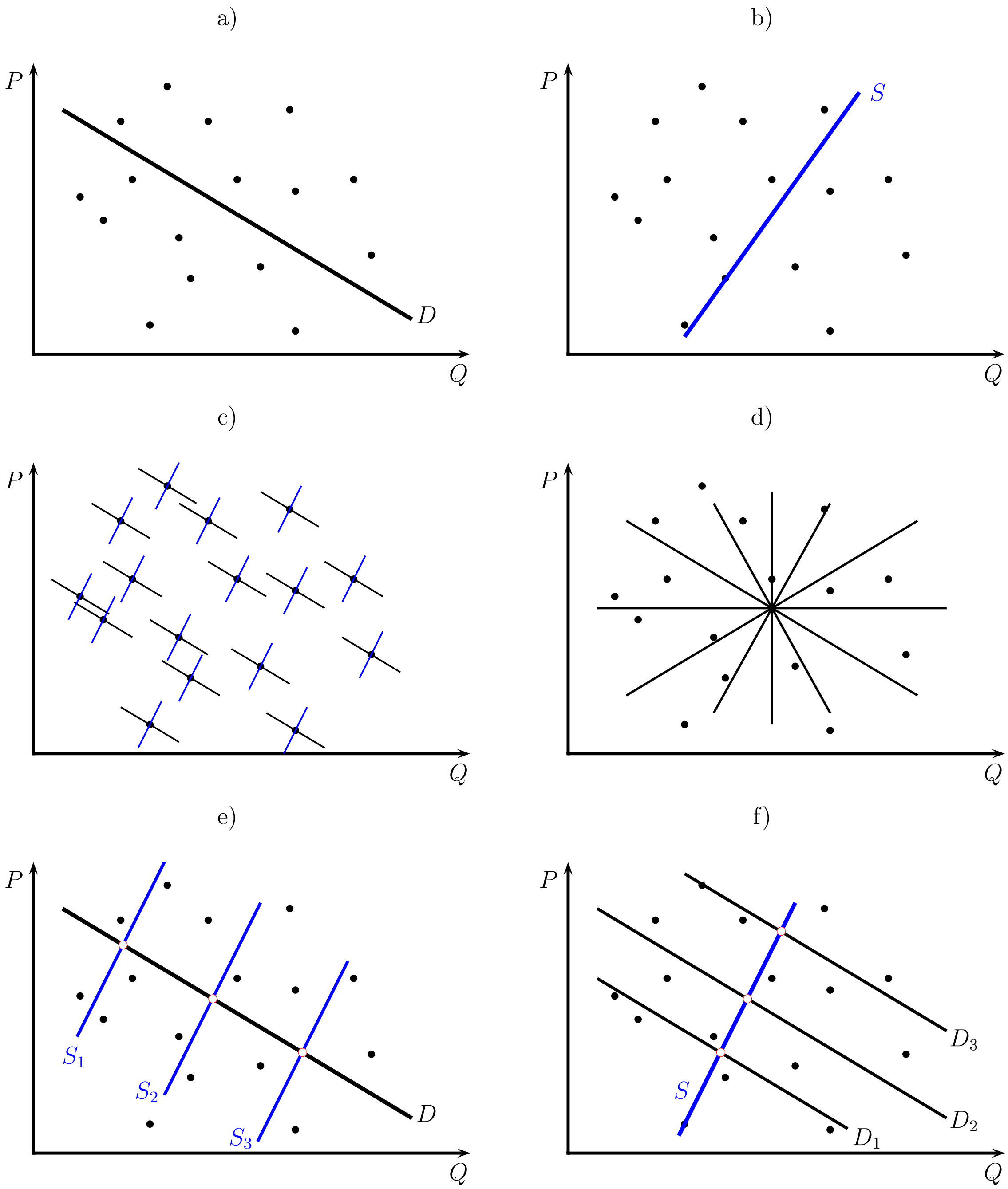


Abbildung 10.13: Das Identifikationsproblem: Wenn wir nur Gleichgewichtspunkte beobachten sind die Koeffizienten ohne zusätzliche Information nicht identifizierbar. Wenn aber eine exogene Variable *nur* das Angebot verschiebt (z.B. das Wetter) wird dadurch die Nachfragefunktion identifizierbar, usw. [lokal, www]

> "For fifty years methods have existed to diagnose whether a conceptual model is unidentifiable, but it appears marketing scholars do not regularly check identification before estimation. To confirm this, all conceptual models published in the Journal of Marketing from 1995 to 1999 are analyzed using the traditional diagnostic methods for identification. . . . Two-thirds of the published conceptual models contain relationships

> that are unidentifiable. These relationships have been empirically estimated, although it is impossible to measure their parameters validly. The published empirical estimates are spurious and cannot be trusted to represent the behavior they claim to measure until the identification problem has been corrected. The theory, not the statistics, must change to validate the measurements, so the paper concludes with suggestions that can help avoid unidentifiable conceptual theories." Hess (2001)

Es sollte aber auch klar sein, dass die Identifikation einer Gleichung nicht erforderlich ist, wenn anstelle der Strukturparameter nur eine Prognose angestrebt wird, da diese unmittelbar mit Hilfe der reduzierten Gleichungen erstellt werden kann (sofern exogene erklärende Variablen existieren!). Wenn aber die Koeffizienten der Strukturgleichungen geschätzt werden sollen muss vorher die Identifikation geprüft werden. Dazu ist offensichtlich zusätzliche Information erforderlich.

Stellen wir uns z.B. vor, die Nachfrage im obigen Modell sei zusätzlich vom Einkommen Y abhängig:

$$\begin{aligned} \text{Angebot:} \quad Q_t &= \alpha_1 + \alpha_2 P_t + u_t \\ \text{Nachfrage:} \quad Q_t &= \beta_1 + \beta_2 P_t + \beta_3 Y_t + v_t \end{aligned}$$

diesem Fall (und wenn Y genügend schwankt) wird die Nachfragekurve im Zeitablauf verschoben, und es ist deshalb möglich eine Angebotskurve zu zeichnen. Dieser Fall ist in Panel f) von Abbildung 10.13 dargestellt.

Durch die zusätzliche Variable in der Nachfragefunktion wird die Angebotsfunktion identifizierbar, obwohl es nach wie vor keine Möglichkeit gibt die Nachfragefunktion zu identifizieren (d.h. das Interzept und den Steigungskoeffizienten der Nachfragefunktion zu schätzen)! Würde hingegen die Angebotskurve z.B. wegen Temperaturschwankungen laufend verschoben, so würde dadurch die Nachfragefunktion identifizierbar (siehe Panel e) von Abbildung 10.13).

Enthalten alle Strukturgleichungen eines simulatanen Gleichungssystems die gleichen Variablen, so kann aus den Daten nicht ermittelt werden, welche Gleichung welche ist, die Strukturkoeffizienten sind nicht ermittelbar. Die Strukturkoeffizienten einer bestimmten Gleichung können nur ermittelt werden, wenn diese Gleichung z.B. bestimmte Variablen *nicht* enthält, damit Schwankungen dieser Variablen die anderen Gleichungen verschieben und somit der Verlauf der in Frage stehenden Strukturgleichung ermittelt werden kann. Man spricht in diesem Zusammenhang von *'exclusion restrictions'*.

Der Kern des Problems besteht darin, dass nicht notwendigerweise eine eindeutige Beziehung zwischen den Parametern der Strukturform und den Parametern der reduzierten Form existiert.

In vielen Fällen, wie z.B. im Nachfrage- Angebotsmodell ohne exogene Variablen, ist die reduzierte Form mit unendlich vielen Strukturformen kompatibel, und es gibt keine Möglichkeit aus den Parametern der reduzierten Form die Parameter der Strukturform zu berechnen.

Wann immer es mehrere Strukturen gibt, die mit den Parametern des reduzierten Modells kompatibel sind, nennt man diese Strukturen '*beobachtungsäquivalent*', d.h. sie können empirisch nicht unterschieden werden.

Eine Struktur ist *'identifizierbar'*, wenn es keine beobachtungsäquivalente Strukturen gibt. Dabei geht es um die Frage, ob aus den Gleichungen der reduzierten Form alle Parameter des Systems in Strukturform bestimmt werden können.

Ein einfaches Beispiel für eine nicht identifizierbare Struktur ist *perfekte Multikollinearität*, in diesem Fall sind unendlich viele Linearkombinationen der Koeffizienten $\boldsymbol{\beta}$ mit den Daten komaptibel, deshalb gibt es keine Möglichkeit einen eindeutigen Koeffizientenvektor $\boldsymbol{\beta}$ zu berechnen. Im Einzelgleichungsmodell ist Nichtidentifizierbarkeit häufig einfach zu erkennen, weil die Schätzer nicht definiert sind, im Mehrgleichungsmodell ist dies leider nicht ganz so einfach.

Generell können hinsichtlich der Identifizierbarkeit drei mögliche Fälle unterschieden werden,

1. es gibt keine Lösung, die Koeffizienten der Strukturform können *nicht* aus den Koeffizienten der reduzierten Form berechnet werden. In diesem Fall sagt man diese Gleichung ist **nicht identifiziert**, bzw. unteridentifiziert.
2. Wenn es eine eindeutige Lösung gibt, wenn alle Koeffizienten einer Gleichung in Strukturform eindeutig aus den reduzierten Gleichungen ermittelt werden können. In diesem Fall ist diese Gleichung **eindeutig** oder *exakt identifiziert*.
3. Schließlich kann es mehr als eine Lösung geben. Wenn für zumindest einen Koeffizienten der Strukturform mehrere Werte möglich sind ist diese Gleichung **überidentifiziert**.

Wie wir gleich sehen werden ist es ohne weiteres möglich, dass in einem Gleichungssystem einige Gleichungen identifizierbar sind und andere nicht.

Überprüfung der Identifizierbarkeit

Zur Überprüfung der Identifizierbarkeit stehen zwei Kriterien zur Verfügung das *Abzähl-* und das *Rangkriterium*.

Das *Abzählkriterium* ist sehr einfach überprüfbar, gibt aber nur eine notwendige Bedingung für Identifizierbarkeit an; das *Rangkriterium* liefert hingegen eine hinreichende Bedingung für Identifizierbarkeit, ist aber deutlich schwieriger zu überprüfen.

Im wesentlichen geht es darum, unter welchen Bedingungen zwischen den Matrizen der Strukturform $\boldsymbol{B}$, $\boldsymbol{\Gamma}$ und der $\boldsymbol{\Pi}$ Matrix der Strukturform eine eindeutige Beziehung besteht. Wir werden im Folgenden die Anwendung dieser beiden Kriterien erläutern, ohne genauer auf die theoretischen Hintergründe einzugehen.

> Das **Abzählkriterium** (*order condition*) besagt, dass eine Gleichung identifizierbar ist, wenn die Anzahl der in einer Gleichung nicht vorkommenden prädeterminierten Variablen größer oder gleich ist wie die Anzahl der in der Gleichung vorkommenden endogenen Variablen minus Eins.

Die Anzahl der in der Gleichung vorkommenden endogenen Variablen umfasst die Variablen links und rechts vom Gleichheitszeichen.

Abzählkriterium:

- wenn $g-1 > K-k$ ist die Gleichung *unter*identifiziert,
- wenn $g-1 = K-k$ ist die Gleichung genau identifiziert,
- wenn $g-1 < K-k$ ist die Gleichung *über*identifiziert.

mit

G: die Anzahl der endogenen Variablen im System
K: die Anzahl der prädeterminierten Variablen im System (inkl. Interzept!)
g: die Anzahl der endogenen Variablen in der betreffenden Gleichung
k: die Anzahl der prädeterminierten Variablen in der Gleichung

Eine äquivalente Form dieses Kriteriums besagt, dass eine notwendige Bedingung für die Identifizierbarkeit einer Gleichung ist, dass *die Anzahl aller in der Gleichung nicht vorkommenden Variablen größer oder gleich sein muss als die Anzahl aller endogenen Variablen (bzw. aller Gleichungen) im Modell minus Eins.*

Wie wir bereits erläutert haben läuft das Argument im Kern darauf hinaus, dass das Strukturmodell *a priori Restriktionen* erfüllen muss, damit eine Gleichung identifizierbar ist. Die häufigste Form sind Nullrestriktionen, d.h., dass bestimmte Parameter in manchen Gleichungen aus theoretischen Gründen den Wert Null haben müssen, also in diesen Gleichung nicht vorkommen dürfen!.

Das Abzählkriterium (*order condition*) ist nur eine notwendige, aber keine hinreichende Bedingung für Identifizierbarkeit. Selbst wenn das Abzählkriterium erfüllt ist kann es nur als Faustregel für die Identifizierbarkeit einer Gleichung gelten. Stellen wir uns z.B. vor, dass im Angebots- Nachfragemodell

$$\begin{aligned} \text{Angebot:} \quad Q_t &= \alpha_1 + \alpha_2 P_t + u_t \\ \text{Nachfrage:} \quad Q_t &= \beta_1 + \beta_2 P_t + \beta_3 Y_t + v_t \end{aligned}$$

das Einkommen Y_t kaum schwankt, bzw. dass der Koeffizient β_3 annähernd Null ist. In diesem Fall ist das Abzählkriterium zwar erfüllt, aber trotzdem wird es nur schwer möglich sein zwischen Nachfrage- und Angebotsfunktion verlässlich zu unterscheiden.

Eine notwendige und hinreichende Bedingung für Identifizierbarkeit ist das **Rang-Kriterium** (*rank condition*), auf welches hier nicht näher eingegangen wird.

10.10 Zusammenfassung

Endogene Regressoren sind mit den Störtermen korreliert, dies führt dazu, dass OLS Schätzer *weder erwartungstreu noch konsistent* sind!

Folgende Fälle führen zu endogenen Regressoren:

- **Fehlende relevante Variablen** (*'omitted variable bias'*)
- **Simultane Kausalität** (*'reverse causality'*, *'omitted equation bias'*)
- **Messfehler in den x Variablen**
- **Selektionsprobleme** (*'selection bias'*)
- **Unbeobachtete Heterogenität** (*'unobserverved heterogeneity'*)

wobei die letzten beiden Fälle häufig als Spezialfälle der ersten beiden Fälle dargestellt werden können, und deshalb hier nicht extra diskutiert werden.

Geeignete Instrumentvariablen sollen ähnlich wie eine Intervention in einem Experiment wirken, d.h. *nur* über den endogenen Regressor auf y einwirken. Technisch ausgedrückt, Instrumentvariablen müssen *relevant* (d.h. $\text{cov}(x, z) \neq 0$, oder besser, möglichst groß), und *exogen* (d.h. $\text{cov}(z, \varepsilon) = 0$) sein!

Tatsächlich ist es äußerst schwierig geeignete Instrumentvariablen zu finden, und 'schwache' Instrumentvariablen führen nicht selten dazu, dass man vom Regen in die Traufe gerät.

Literaturverzeichnis

Angrist, J. D. and Pischke, J.-S. (2008), *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press.

Cameron, A. C. and Trivedi, P. K. (2005), *Microeconometrics: Methods and Applications*, Cambridge University Press.

Davidson, R. and MacKinnon, J. G. (1989), 'Testing for consistency using artificial regressions', *Econometric Theory* **5**(3), 363–384.
URL: *http://www.jstor.org/stable/3532374*

Deaton, A. and Cartwright, N. (2016), Understanding and misunderstanding randomized controlled trials, Working Paper 22595, National Bureau of Economic Research.

Graddy, K. (1995), 'Testing for imperfect competition at the fulton fish market', *The RAND Journal of Economics* **26**(1), pp. 75–92.

Graddy, K. (2006), 'Markets: The fulton fish market', *The Journal of Economic Perspectives* **20**(2), pp. 207–220.

Graddy, K. and Kennedy, P. (2010), 'When are supply and demand determined recursively rather than simultaneously?', *Eastern Economic Journal* **36**(2), pp. 188–197.

Greene, W. (2003), *Solutions Manual: Econometric Analysis*, 5^{th} edn, Pearson Education.

Hausman, J. A. (1978), 'Specification tests in econometrics', *Econometrica* **46**(6), 1251–1271.
URL: *http://www.jstor.org/stable/1913827*

Hess, J. (2001), 'Unidentifiable relationships in conceptual marketing models', *Review of Marketing Science* **WP No. 316**.

Hume, D. (1993), *Eine Untersuchung über den menschlichen Verstand*, 12 edn, Meiner Verlag. Hrsg. von Jens Kulenkampff, übersetzt von Raoul Richter.

Imbens, G. W. (2014), Instrumental Variables: An Econometrician's Perspective, Working Paper 19983, National Bureau of Economic Research.
URL: *http://www.nber.org/papers/w19983*

Leamer, E. E. (1985), 'Vector autoregressions for causal inference?', *Carnegie-Rochester Conference Series on Public Policy* **22**, 255 – 304.

Lewis, D. (1973), 'Causation', *The Journal of Philosophy* **70**(17), 556–567.

Pearl, J. (2000), *Causality: Models, Reasoning, and Inference*, reprinted with corrections edn, Cambridge University Press.

Salsburg, D. (2002), *The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century*, 1st edition edn, Holt Paperbacks.
URL: *http://amazon.com/o/ASIN/0805071342/*

Sargan, D. (1988), *Lectures on Advanced Econometric Theory*, Blackwell Publishers.

Stock, J. H. and Watson, M. W. (2006), *Introduction to Econometrics*, 2 edn, Addison Wesley.

Woodward, J. (2016), Causation and Manipulability, *in* E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', winter 2016 edn, Metaphysics Research Lab, Stanford University.