



Endogenität, Kausalität, oder das große “***Was wäre wenn . . .***”

Grundlagen der Ökonometrie

herbert.stocker@uibk.ac.at

www.hsto.info/econometrics

Kausalität

- Regressionen sind eine Methode für *bedingte Mittel- bzw. Erwartungswertvergleiche*

Kausalität

- Regressionen sind eine Methode für *bedingte Mittel- bzw. Erwartungswertvergleiche*
- Regressionen erlauben per se *keine* Beurteilung von Ursachen-Wirkungszusammenhängen!!!
Aber: Suche nach Ursachen und Wirkungen bilden seit jeher den Kern unserer Wissenschaft, z.B. Adam Smith (1776): "*An Inquiry into the Nature and Causes of the Wealth of Nations*"

Kausalität

- Regressionen sind eine Methode für *bedingte Mittel- bzw. Erwartungswertvergleiche*
- Regressionen erlauben per se *keine* Beurteilung von Ursachen-Wirkungszusammenhängen!!!
Aber: Suche nach Ursachen und Wirkungen bilden seit jeher den Kern unserer Wissenschaft, z.B. Adam Smith (1776): *“An Inquiry into the Nature and **Causes** of the Wealth of Nations”*
- *“When we study nonexperimental data, correlation is in the data while causation is in the mind or heart of the observer. That’s where considered judgments and wishful thinking originate.”* (Leamer, 2009)
 - Leben regelmäßige Kirchgänger länger?
 - Leben Anhänger der Kreta-Diät länger?
 - Haben Länder mit unabhängigen Notenbanken niedrigere Inflationsraten (und höheres Wachstum)?
 - ...

Kausalität

- David Hume (1711 – 1776): wir können *Kausalität* nicht sinnlich wahrnehmen, nur eine Abfolge von Ereignissen

“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.”

[zitiert nach <https://de.wikipedia.org/wiki/Kausalität>].

Kausalität

- David Hume (1711 – 1776): wir können *Kausalität* nicht sinnlich wahrnehmen, nur eine Abfolge von Ereignissen

“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.”

[zitiert nach <https://de.wikipedia.org/wiki/Kausalität>].

⇒ Erfahrung der *Regularität* (*‘constant conjunction’*) wird als Ursachen und Wirkung *erlebt!*

Kausalität

David Hume: zwei zentrale Ideen:

- ① **Zeitliche Anordnung:** die Wirkung kann nicht vor der Ursache eintreten (asymmetrisch!)

"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." \Rightarrow Granger-Kausalität

Kausalität

David Hume: zwei zentrale Ideen:

- ① **Zeitliche Anordnung:** die Wirkung kann nicht vor der Ursache eintreten (asymmetrisch!)

"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." \Rightarrow Granger-Kausalität

- ② **Kontrafaktische Situation:** was wäre passiert, wenn das verursachende Ereignis *nicht* eingetreten wäre?

"We may define a cause to be an object, where, if the first object had not been, the second never had existed."

Kausalität

David Hume: zwei zentrale Ideen:

- ① **Zeitliche Anordnung:** die Wirkung kann nicht vor der Ursache eintreten (asymmetrisch!)

"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." ⇒ Granger-Kausalität

- ② **Kontrafaktische Situation:** was wäre passiert, wenn das verursachende Ereignis *nicht* eingetreten wäre?

"We may define a cause to be an object, where, if the first object had not been, the second never had existed."

→ Kontrafaktische Denken entspricht ökonomischer Denkweise (z.B. Alternativkosten)!

Kausalität

David Hume: zwei zentrale Ideen:

- ① **Zeitliche Anordnung:** die Wirkung kann nicht vor der Ursache eintreten (asymmetrisch!)

"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." ⇒ Granger-Kausalität

- ② **Kontrafaktische Situation:** was wäre passiert, wenn das verursachende Ereignis *nicht* eingetreten wäre?

"We may define a cause to be an object, where, if the first object had not been, the second never had existed."

→ Kontrafaktische Denken entspricht ökonomischer Denkweise (z.B. Alternativkosten)!

"Treffen sich zwei Ökonomen, fragt der eine: 'Wie geht es Ihnen, Herr Kollege', antwortet der andere: 'Verglichen mit was?'"

Kausalität

Kontrafaktische Situation:

⇒ **Problem:** wir können nie faktische und kontrafaktische Situation gleichzeitig beobachten!

Kausalität

Kontrafaktische Situation:

⇒ **Problem:** wir können nie faktische und kontrafaktische Situation gleichzeitig beobachten!

- Wären Personen später gestorben, wenn Sie in die Kirche gegangen wären?

Kausalität

Kontrafaktische Situation:

⇒ **Problem:** wir können nie faktische und kontrafaktische Situation gleichzeitig beobachten!

- Wären Personen später gestorben, wenn Sie in die Kirche gegangen wären?
- Wären Personen später gestorben, wenn Sie sich nach der Kreta Diät ernährt hätten?

Kausalität

Kontrafaktische Situation:

⇒ **Problem:** wir können nie faktische und kontrafaktische Situation gleichzeitig beobachten!

- Wären Personen später gestorben, wenn Sie in die Kirche gegangen wären?
- Wären Personen später gestorben, wenn Sie sich nach der Kreta Diät ernährt hätten?
- Wäre die Inflationsrate niedriger, wenn die Nationalbank unabhängiger wäre?

Kausalität

Kontrafaktische Situation:

⇒ **Problem:** wir können nie faktische und kontrafaktische Situation gleichzeitig beobachten!

- Wären Personen später gestorben, wenn Sie in die Kirche gegangen wären?
- Wären Personen später gestorben, wenn Sie sich nach der Kreta Diät ernährt hätten?
- Wäre die Inflationsrate niedriger, wenn die Nationalbank unabhängiger wäre?
- Hinter jeder Kausalfrage steht explizit oder implizit eine Modellvorstellung:
Was wäre passiert wenn . . .

Kausalität

Kontrafaktische Situation:

⇒ **Problem:** wir können nie faktische und kontrafaktische Situation gleichzeitig beobachten!

- Wären Personen später gestorben, wenn Sie in die Kirche gegangen wären?
- Wären Personen später gestorben, wenn Sie sich nach der Kreta Diät ernährt hätten?
- Wäre die Inflationsrate niedriger, wenn die Nationalbank unabhängiger wäre?
- Hinter jeder Kausalfrage steht explizit oder implizit eine Modellvorstellung:
Was wäre passiert wenn . . .
- **Kausalfragen sind immer entscheidend, wenn die Wirksamkeit eines Eingriffs in ein System beurteilt werden soll!**

→ Medizin, → Sozialwissenschaften

Kausalität

Kontrafaktische Situation:

⇒ **Problem:** wir können nie faktische und kontrafaktische Situation gleichzeitig beobachten!

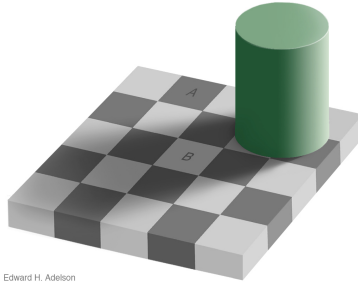
- Wären Personen später gestorben, wenn Sie in die Kirche gegangen wären?
- Wären Personen später gestorben, wenn Sie sich nach der Kreta Diät ernährt hätten?
- Wäre die Inflationsrate niedriger, wenn die Nationalbank unabhängiger wäre?
- Hinter jeder Kausalfrage steht explizit oder implizit eine Modellvorstellung:
Was wäre passiert wenn . . .
- **Kausalfragen sind immer entscheidend, wenn die Wirksamkeit eines Eingriffs in ein System beurteilt werden soll!**

→ Medizin, → Sozialwissenschaften

⇒ Unterscheidung zwischen (Schein-)Korrelation und Ursachen-Wirkungszusammenhang!!!

Kausalität

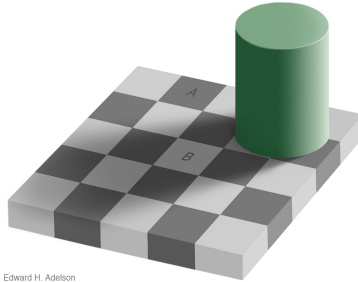
- Unsere Wahrnehmung ist nicht *objektiv*: A und B sind gleich hell!



Edward H. Adelson

Kausalität

- Unsere Wahrnehmung ist nicht *objektiv*: A und B sind gleich hell!

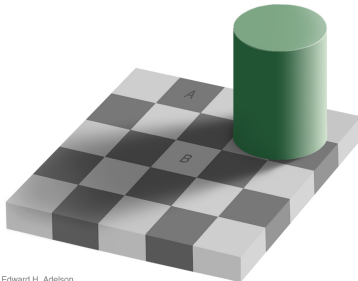


Edward H. Adelson

- Evolution: es war wichtig, sich im Schatten versteckende Gefahren zu erkennen!

Kausalität

- Unsere Wahrnehmung ist nicht *objektiv*: A und B sind gleich hell!

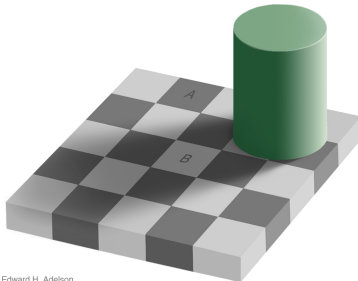


Edward H. Adelson

- Evolution: es war wichtig, sich im Schatten versteckende Gefahren zu erkennen!
- Ähnlich: unser schnelles, intuitives Denken (Kahnemann: Typ I Denken) *projiziert* Zusammenhänge; vgl. Autovervollständigung!

Kausalität

- Unsere Wahrnehmung ist nicht *objektiv*: A und B sind gleich hell!

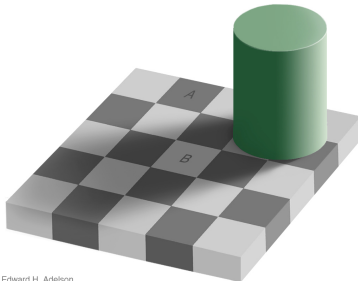


Edward H. Adelson

- Evolution: es war wichtig, sich im Schatten versteckende Gefahren zu erkennen!
- Ähnlich: unser schnelles, intuitives Denken (Kahnemann: Typ I Denken) *projiziert* Zusammenhänge; vgl. Autovervollständigung!
- funktioniert oft gut ...

Kausalität

- Unsere Wahrnehmung ist nicht *objektiv*: A und B sind gleich hell!



Edward H. Adelson

- Evolution: es war wichtig, sich im Schatten versteckende Gefahren zu erkennen!
- Ähnlich: unser schnelles, intuitives Denken (Kahnemann: Typ I Denken) *projiziert* Zusammenhänge; vgl. Autovervollständigung!
- funktioniert oft gut ...
aber: Ideologien, Aberglaube, Vorurteile, Verschwörungstheorien ...

Kausalität

- Zentral hinter den meisten modernen Kausalitätstheorien ist die Vorstellung eines verursachenden Ereignisses, einer *Intervention*, Manipulation ...

Kausalität

- Zentral hinter den meisten modernen Kausalitätstheorien ist die Vorstellung eines verursachenden Ereignisses, einer *Intervention*, Manipulation . . .
- z.B.: “Hat die Tablette den Kopfschmerz beseitigt, oder wäre der Kopfschmerz auch ohne Tablette weggegangen?”

“Haben die Werbeausgaben die Umsätze erhöht, oder haben höhere Umsätze zu höheren Werbeausgaben geführt?”

Kausalität

- Zentral hinter den meisten modernen Kausalitätstheorien ist die Vorstellung eines verursachenden Ereignisses, einer *Intervention*, Manipulation . . .
- z.B.: “Hat die Tablette den Kopfschmerz beseitigt, oder wäre der Kopfschmerz auch ohne Tablette weggegangen?”
“Haben die Werbeausgaben die Umsätze erhöht, oder haben höhere Umsätze zu höheren Werbeausgaben geführt?”
- Hinter dem Kausalitätsgedanken steht nicht nur die Frage ‘ob’, sondern v.a. die Frage eines **‘warum’**!

Kausalität

- Zentral hinter den meisten modernen Kausalitätstheorien ist die Vorstellung eines verursachenden Ereignisses, einer *Intervention*, Manipulation . . .
- z.B.: “Hat die Tablette den Kopfschmerz beseitigt, oder wäre der Kopfschmerz auch ohne Tablette weggegangen?”
“Haben die Werbeausgaben die Umsätze erhöht, oder haben höhere Umsätze zu höheren Werbeausgaben geführt?”
- Hinter dem Kausalitätsgedanken steht nicht nur die Frage ‘ob’, sondern v.a. die Frage eines **‘warum’**!
- Zwei Möglichkeiten:

Kausalität

- Zentral hinter den meisten modernen Kausalitätstheorien ist die Vorstellung eines verursachenden Ereignisses, einer *Intervention*, Manipulation . . .
- z.B.: “Hat die Tablette den Kopfschmerz beseitigt, oder wäre der Kopfschmerz auch ohne Tablette weggegangen?”
“Haben die Werbeausgaben die Umsätze erhöht, oder haben höhere Umsätze zu höheren Werbeausgaben geführt?”
- Hinter dem Kausalitätsgedanken steht nicht nur die Frage ‘ob’, sondern v.a. die Frage eines **‘warum’**!
- Zwei Möglichkeiten:
 - ‘Nachbau’ der Realität in theoretischen Modellen und Simulation der kontrafaktischen Situation (z.B. Außenhandelsmodelle)

Kausalität

- Zentral hinter den meisten modernen Kausalitätstheorien ist die Vorstellung eines verursachenden Ereignisses, einer *Intervention*, Manipulation . . .
- z.B.: “Hat die Tablette den Kopfschmerz beseitigt, oder wäre der Kopfschmerz auch ohne Tablette weggegangen?”
“Haben die Werbeausgaben die Umsätze erhöht, oder haben höhere Umsätze zu höheren Werbeausgaben geführt?”
- Hinter dem Kausalitätsgedanken steht nicht nur die Frage ‘ob’, sondern v.a. die Frage eines **‘warum’**!
- Zwei Möglichkeiten:
 - ‘Nachbau’ der Realität in theoretischen Modellen und Simulation der kontrafaktischen Situation (z.B. Außenhandelsmodelle)
 - (Naturwissenschaftliche) Experimente

Kausalität

Experimente: Galileo Galilei (1564 – 1642): Versuche auf der schiefen Ebene



Kausalität

Experimente: Galileo Galilei (1564 – 1642): Versuche auf der schiefen Ebene



① Kontrolle durch Herstellung von *Laborbedingungen*;

Kausalität

Experimente: Galileo Galilei (1564 – 1642): Versuche auf der schiefen Ebene



- ① **Kontrolle** durch Herstellung von *Laborbedingungen*;
- ② **Intervention:** Experimentator greift gezielt ein und manipuliert (üblicherweise) einen einzelnen isolierten Einflussfaktor.

Kausalität

Experimente:

Da alle anderen möglichen Einflussfaktoren konstant gehalten werden kann die beobachtete Wirkung unmittelbar auf die Intervention des Experimentators zurückgeführt werden.

Kontrafaktische Interpretation: hätte die Intervention des Experimentators nicht stattgefunden, hätte die Wirkung nicht beobachtet werden können.

Kausalität

Experimente:

Da alle anderen möglichen Einflussfaktoren konstant gehalten werden kann die beobachtete Wirkung unmittelbar auf die Intervention des Experimentators zurückgeführt werden.

Kontrafaktische Interpretation: hätte die Intervention des Experimentators nicht stattgefunden, hätte die Wirkung nicht beobachtet werden können.

Wenn perfekte Kontrolle unmöglich ist:

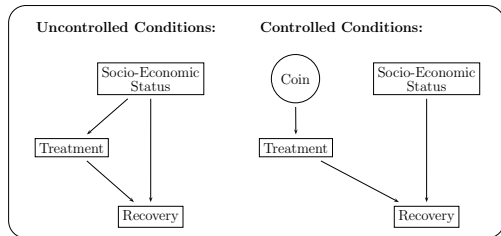
⇒ **Randomisierung** (*Randomized Controlled Trials, RCT*)

Bei perfekter Randomisierung (*random assignment*) sollten keine *systematischen* Unterschiede zwischen Interventions- und Kontrollgruppe auftreten.

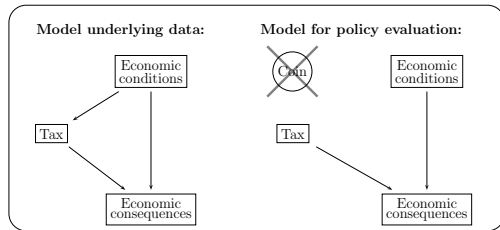
Kausalität

Kontrollierte vs. nicht kontrollierte Experimente:

Example 1: Controlled Experimentation



Example 2: Policy Analysis



J. Pearl, Causality

Empirische Studien

- **Experimente:** Goldstandard? Nicht immer durchführbar!
siehe z.B. A. Deaton and N. Cartwright (2016), *Understanding and Misunderstanding Randomized Controlled Trials*, NBER Working Paper 22595

Empirische Studien

- **Experimente:** Goldstandard? Nicht immer durchführbar!
siehe z.B. A. Deaton and N. Cartwright (2016), *Understanding and Misunderstanding Randomized Controlled Trials*, NBER Working Paper 22595
- **Beobachtungsstudien:**
 - weder Kontrolle noch gezielte Intervention.
 - erlauben per se keine Kausalaussagen!!!

Empirische Studien

- **Experimente:** Goldstandard? Nicht immer durchführbar!
siehe z.B. A. Deaton and N. Cartwright (2016), *Understanding and Misunderstanding Randomized Controlled Trials*, NBER Working Paper 22595
- **Beobachtungsstudien:**
 - weder Kontrolle noch gezielte Intervention.
 - erlauben per se keine Kausalaussagen!!!

Anforderungen an (empirische) Studien:

- **Reliabilität:** Reproduzierbarkeit von Ergebnissen unter gleichen Bedingungen?
- **Validität:** ('Gültigkeit' der Meßergebnisse)
 - *Interne Validität:* wird tatsächlich das gemessen, was wir messen wollen? Ist diese Schätzung konsistent?
 - *Externe Validität:* Sind Ergebnisse z.B. auch auf andere Grundgesamtheiten übertragbar?

Endogenität

Endogene Regressoren (d.h. $E(\varepsilon_i|\mathbf{X}) \neq 0$) gefährden interne Validität!

z.B.

- Fehlende relevante Variablen (*'omitted variables'*); mit den Spezialfällen Selektionsprobleme (*'selection bias'*) und Unbeobachtete Heterogenität (*'unobserved heterogeneity'*)
- Simultane Kausalität (*'simultaneous causality bias'*)
- Messfehler in den x Variablen (*'errors-in-variables bias'*);

(werden noch einzeln diskutiert!)

Endogenität

Endogene Regressoren (d.h. $E(\varepsilon_i|\mathbf{X}) \neq 0$) gefährden interne Validität!

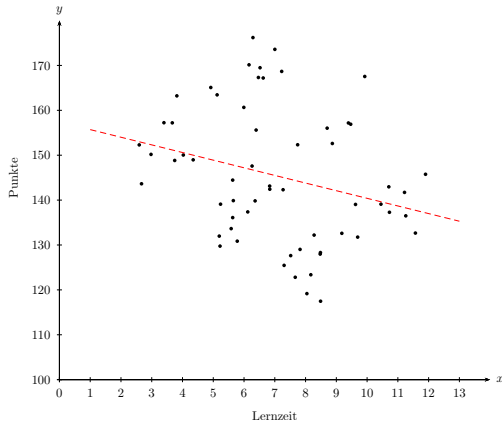
z.B.

- Fehlende relevante Variablen (*'omitted variables'*); mit den Spezialfällen Selektionsprobleme (*'selection bias'*) und Unbeobachtete Heterogenität (*unobserved heterogeneity*)
- Simultane Kausalität (*'simultaneous causality bias'*)
- Messfehler in den x Variablen (*'errors-in-variables bias'*);

(werden noch einzeln diskutiert!)

Achtung: die Begriffe *Endogenität* und *Exogenität* werden in Ökonomik und Ökonometrie nicht deckungsgleich gebraucht!

Beispiel: Lernzeit (LZ) und Lernerfolg (Pkt)



$$\text{Pkt} = 157.404 - 1.701 \text{ LZ}$$

$(6.004)^{***} \quad (0.807)^{**}$

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

Was passiert hier???

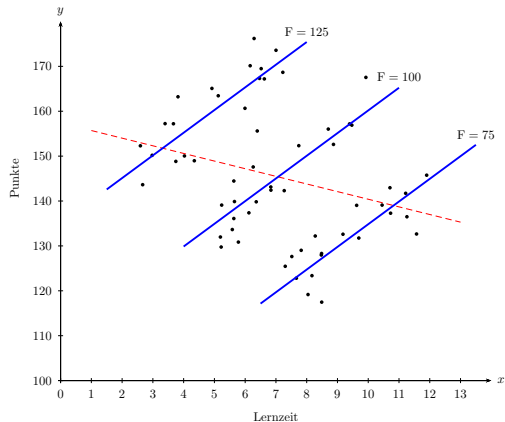
Ein Beispiel

Fehlspezifikation! unbeobachtete Fähigkeiten (F)!!!

‘Wahres’ Modell (DGP):

$$\text{Pkt}_i = 10 + 5\text{LZ}_i + 1F_i + \varepsilon_i$$

$$\text{mit LZ}_i = 12 - 0.1F_i + v_i$$



Ein Beispiel

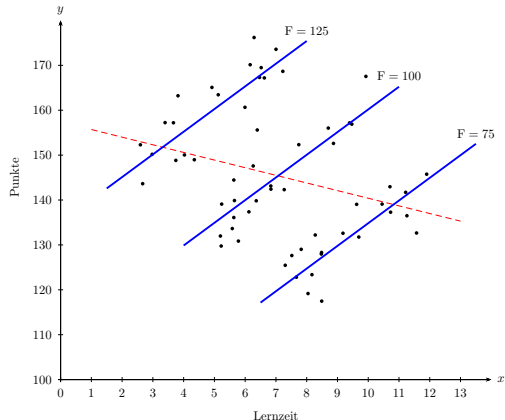
Fehlspezifikation! unbeobachtete Fähigkeiten (F)!!!

‘Wahres’ Modell (DGP):

$$\text{Pkt}_i = 10 + 5\text{LZ}_i + 1\text{F}_i + \varepsilon_i$$

$$\text{mit LZ}_i = 12 - 0.1\text{F}_i + v_i$$

Klassischer ‘*omitted variable bias*’!!!



Omitted Variable Bias

Zur Erinnerung: “wahres” Modell (mit $\ddot{x}_i := x_i - \bar{x}$ etc.) sei

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

‘Falsches’ (kurzes) Modell:

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

Der geschätzte OLS Koeffizient des ‘falschen’ Modells (mit Einsetzen der PRF) ist

$$\hat{\beta}_2^* = \frac{\sum \ddot{x}_{i2} \ddot{y}_i}{\sum \ddot{x}_{i2}^2} = \frac{\sum \ddot{x}_{i2} \overbrace{(\beta_2 \ddot{x}_{i2} + \beta_3 \ddot{x}_{i3} + \varepsilon_i)}^{\ddot{y}_i}}{\sum \ddot{x}_{i2}^2}$$

Erwartungstreue:

$$\mathbb{E}(\hat{\beta}_2^*) = \mathbb{E}(\beta_2) + \mathbb{E}\left(\beta_3 \frac{\sum \ddot{x}_{i2} \ddot{x}_{i3}}{\sum \ddot{x}_{i2}^2}\right) + \mathbb{E}\left(\frac{\sum \ddot{x}_{i2} \varepsilon_i}{\sum \ddot{x}_{i2}^2}\right)$$

Omitted Variable Bias

$$\begin{aligned}\mathbb{E}(\hat{\beta}_2^*) &= \mathbb{E}(\beta_2) + \mathbb{E}\left(\beta_3 \frac{\sum \ddot{x}_{i2}\ddot{x}_{i3}}{\sum \ddot{x}_{i2}^2}\right) + \mathbb{E}\left(\frac{\sum \ddot{x}_{i2}\varepsilon_i}{\sum \ddot{x}_{i2}^2}\right) \\ &= \beta_2 + \beta_3 \frac{\sum \ddot{x}_{i2}\ddot{x}_{i3}}{\sum \ddot{x}_{i2}^2} \\ &= \beta_2 + \beta_3 \frac{\text{cov}(x_2, x_3)}{\text{var}(x_2)} = \beta_2 + \beta_3 \delta\end{aligned}$$

mit $x_{i3} = \delta_1 + \delta_2 x_{i2} + v_i$

Omitted Variable Bias

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

mit $x_{i3} = \delta_1 + \delta_2 x_{i2} + v_i$

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

Omitted Variable Bias (OVB)

Warum führt OVB zu $E(\varepsilon_i|\mathbf{X}) \neq 0$?

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

Im 'kurzen' Modell steckt F im Störterm ε^* , also ist ε^* mit F und LZ korreliert!

Omitted Variable Bias (OVB)

Warum führt OVB zu $E(\varepsilon_i|\mathbf{X}) \neq 0$?

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

Im 'kurzen' Modell steckt F im Störterm ε^* , also ist ε^* mit F und LZ korreliert!

Wenn Störterm Funktion von x :

$$y = \beta_1 + \beta_2 x + \varepsilon(x) \quad \Rightarrow \quad \frac{dy}{dx} = \beta_2 + \frac{d\varepsilon}{dx} := \beta_2^*$$

Omitted Variable Bias (OVB)

Warum führt OVB zu $E(\varepsilon_i|\mathbf{X}) \neq 0$?

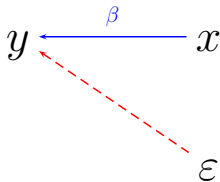
$$\text{Pkt} = \beta_1 + \beta_2 \text{LZ} + \underbrace{\beta_3 F + \varepsilon}_{\varepsilon^*}$$

Im 'kurzen' Modell steckt F im Störterm ε^* , also ist ε^* mit F und LZ korreliert!

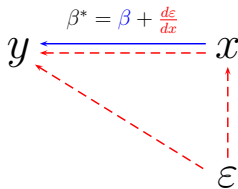
Wenn Störterm Funktion von x :

$$y = \beta_1 + \beta_2 x + \varepsilon(x) \quad \Rightarrow \quad \frac{dy}{dx} = \beta_2 + \frac{d\varepsilon}{dx} := \beta_2^*$$

$$E(\varepsilon|x) = 0:$$



$$E(\varepsilon|x) \neq 0:$$



Omitted Variable Bias (OVB)

Im linearen Fall: DGP: $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$

Interessierender Parameter: β_2

Unbeobachtbar: x_3 , mit $x_{i3} = \delta_1 + \delta_2 x_{i2} + v_i$

Omitted Variable Bias (OVB)

Im linearen Fall: DGP: $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$

Interessierender Parameter: β_2

Unbeobachtbar: x_3 , mit $x_{i3} = \delta_1 + \delta_2 x_{i2} + v_i$

Einsetzen

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

und der marginale Effekt ist

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

Omitted Variable Bias (OVB)

Im linearen Fall: DGP: $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$

Interessierender Parameter: β_2

Unbeobachtbar: x_3 , mit $x_{i3} = \delta_1 + \delta_2 x_{i2} + v_i$

Einsetzen

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

und der marginale Effekt ist

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

$\Rightarrow \beta_2$ **ist nicht identifizierbar!**

Simultane Kausalität

⇒ führt zu ähnlichen Problemen! Warum?

Simultane Kausalität

⇒ führt zu ähnlichen Problemen! Warum?

Lehrbuchbeispiel (Keynes):

$$C_i = 60 + 0.5Y_i + \varepsilon_i \quad \text{mit } \varepsilon_i \sim U(-30, +30)$$

$$Y_i = C_i + I_i + G_i \quad \text{mit } I_i \sim U(30, 80)$$

Simultane Kausalität

⇒ führt zu ähnlichen Problemen! Warum?

Lehrbuchbeispiel (Keynes):

$$C_i = 60 + 0.5Y_i + \varepsilon_i \quad \text{mit } \varepsilon_i \sim U(-30, +30)$$

$$Y_i = C_i + I_i + G_i \quad \text{mit } I_i \sim U(30, 80)$$

Wahres $\beta_2 = 0.5$, aber Monte Carlo Simulation liefert

$$C = -9.175 + 0.808 Y$$

(5.685) (0.024)^{***}

$$R^2 = 0.934, \quad n = 80$$

Was ist passiert?

Simultane Kausalität

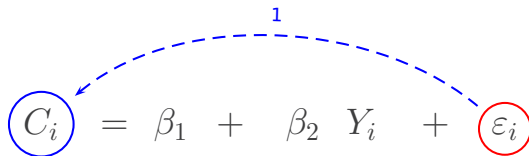
Einfaches keynesianisches Einnahmen-Ausgaben Modell:

$$C_i = \beta_1 + \beta_2 Y_i + \varepsilon_i$$

$$Y_i = C_i + I_i + G_i$$

Simultane Kausalität

Einfaches keynesianisches Einnahmen-Ausgaben Modell:

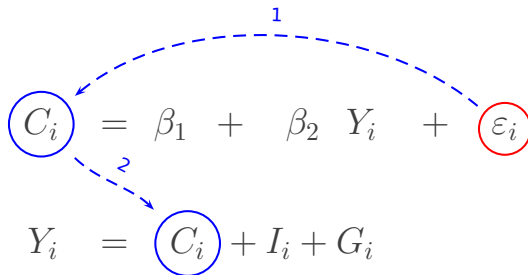
$$\textcircled{C_i} = \beta_1 + \beta_2 Y_i + \textcircled{\varepsilon_i}$$


A diagram illustrating a causal relationship. A blue dashed curved arrow points from the error term ε_i (enclosed in a red circle) to the consumption variable C_i (enclosed in a blue circle). Above the arrow is a blue number '1'.

$$Y_i = C_i + I_i + G_i$$

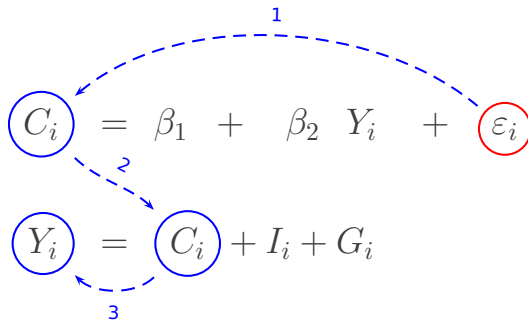
Simultane Kausalität

Einfaches keynesianisches Einnahmen-Ausgaben Modell:



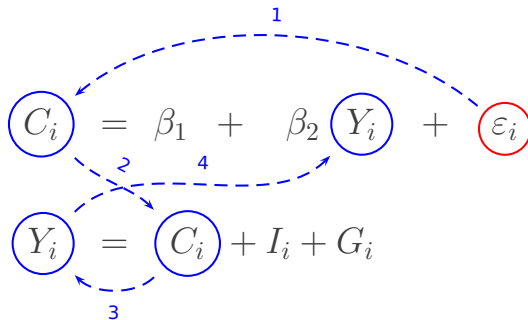
Simultane Kausalität

Einfaches keynesianisches Einnahmen-Ausgaben Modell:



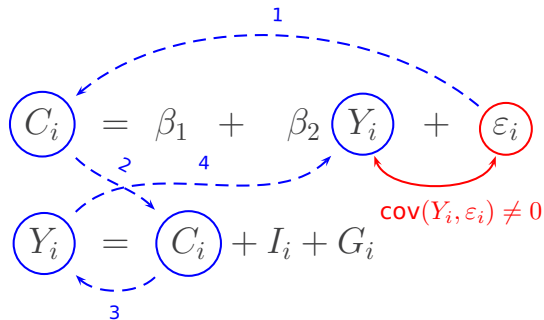
Simultane Kausalität

Einfaches keynesianisches Einnahmen-Ausgaben Modell:



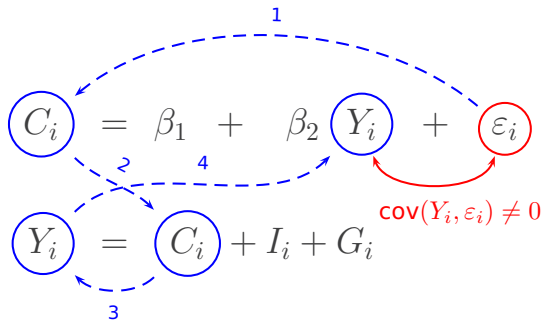
Simultane Kausalität

Einfaches keynesianisches Einnahmen-Ausgaben Modell:



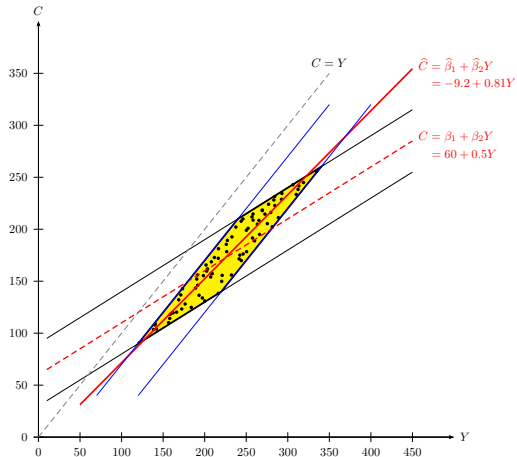
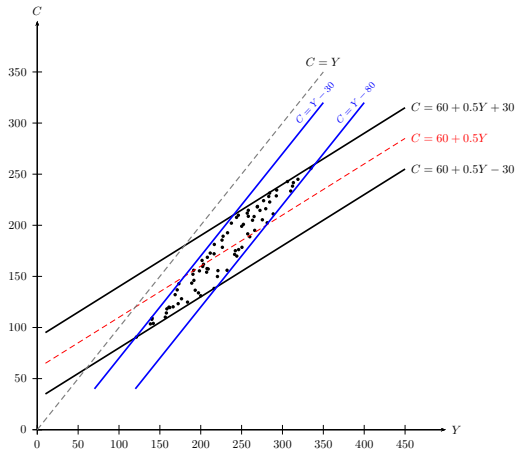
Simultane Kausalität

Einfaches keynesianisches Einnahmen-Ausgaben Modell:

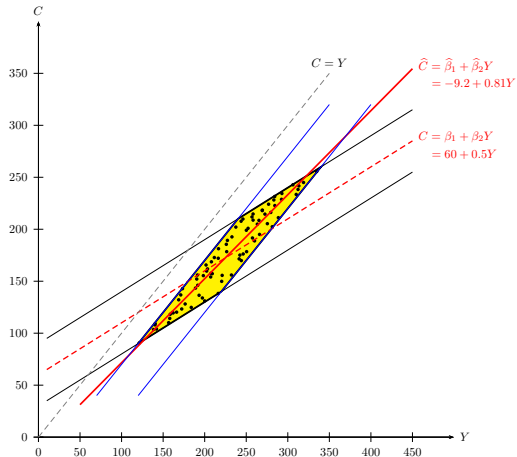
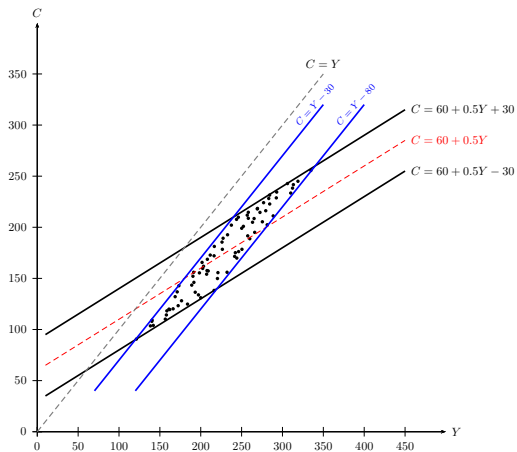


Weil $\text{cov}(Y_i, \varepsilon_i) \neq 0$ ist eine OLS Schätzung der Konsumfunktion weder erwartungstreu noch konsistent!!!

Simultane Kausalität



Simultane Kausalität



⇒ **OLS verzerrt und nicht konsistent!**

Simultane Kausalität

- Schätzung der *reduzierten Form* mit OLS möglich
- Lösung der ‘Strukturform’ nach endogenen Variablen: z.B.
Strukturform:

$$\begin{aligned}C_i &= \beta_1 + \beta_2 Y_i + \varepsilon_i \\Y_i &= C_i + I_i + G_i\end{aligned}$$

Reduzierte Form:

$$\begin{aligned}C_i^* &= \frac{1}{1 - \beta_2} (\beta_1 + \beta_2(I_i + G_i) + \varepsilon_i) \\Y_i^* &= C_i^* + I_i + G_i\end{aligned}$$

- Interesse gilt meist den Parametern der Strukturform!
- In komplexen Modellen kann häufig aus den Schätzungen der reduzierten Form nicht auf die Parameter der Strukturform geschlossen werden
 \Rightarrow **Identifikationsproblem!**

Simultane Kausalität

Wann liegt Interdependenz vor?

Zumindest wenn zur Beschreibung eines DGP mehr als eine interdependente Gleichung benötigt wird! (*omitted equation bias*)

z.B. Angebots- und Nachfragefunktion: beschreiben Verhalten unterschiedlicher Akteure!

Messfehler in den erklärenden Variablen

- Wir interessieren uns für β_2 der PRF $\ddot{y}_i = \beta_2 \ddot{x}_i + \varepsilon_i$ (mit $\ddot{x}_i := x_i - \bar{x}$)
- aber \ddot{x}_i ist nicht beobachtbar, sondern nur ein fehlerhaft gemessenes $\ddot{x}_i^* = \ddot{x}_i + v_i$ mit $v_i \sim \text{i.i.d.}(0, \sigma_{v_i}^2)$ ($v \dots$ epsilon)

Messfehler in den erklärenden Variablen

- Wir interessieren uns für β_2 der PRF $\ddot{y}_i = \beta_2 \ddot{x}_i + \varepsilon_i$ (mit $\ddot{x}_i := x_i - \bar{x}$)
- aber \ddot{x}_i ist nicht beobachtbar, sondern nur ein fehlerhaft gemessenes $\ddot{x}_i^* = \ddot{x}_i + v_i$ mit $v_i \sim \text{i.i.d.}(0, \sigma_{v_i}^2)$ (v ... epsilon)
- Schätzung von $\ddot{y}_i = \beta_2 \ddot{x}_i^* + \varepsilon_i^*$ liefert verzerrtes Ergebnis!

$$\begin{aligned}\ddot{y}_i &= \beta_2 \ddot{x}_i + \varepsilon_i \\ &= \beta_2 (\ddot{x}_i^* - v_i) + \varepsilon_i \\ &= \beta_2 \ddot{x}_i^* + (\varepsilon_i - \beta_2 v_i) \\ \ddot{y}_i &= \beta_2 \ddot{x}_i^* + \varepsilon_i^* = \beta_2 \underbrace{(\ddot{x}_i + v_i)}_{\ddot{x}_i^*} + \underbrace{(\varepsilon_i - \beta_2 v_i)}_{\varepsilon_i^*}\end{aligned}$$

Messfehler in den erklärenden Variablen

- Wir interessieren uns für β_2 der PRF $\ddot{y}_i = \beta_2 \ddot{x}_i + \varepsilon_i$ (mit $\ddot{x}_i := x_i - \bar{x}$)
- aber \ddot{x}_i ist nicht beobachtbar, sondern nur ein fehlerhaft gemessenes $\ddot{x}_i^* = \ddot{x}_i + v_i$ mit $v_i \sim \text{i.i.d.}(0, \sigma_{v_i}^2)$ ($v \dots$ epsilon)
- Schätzung von $\ddot{y}_i = \beta_2 \ddot{x}_i^* + \varepsilon_i^*$ liefert verzerrtes Ergebnis!

$$\begin{aligned}\ddot{y}_i &= \beta_2 \ddot{x}_i + \varepsilon_i \\ &= \beta_2 (\ddot{x}_i^* - v_i) + \varepsilon_i \\ &= \beta_2 \ddot{x}_i^* + (\varepsilon_i - \beta_2 v_i) \\ \ddot{y}_i &= \beta_2 \ddot{x}_i^* + \varepsilon_i^* = \beta_2 \underbrace{(\ddot{x}_i + v_i)}_{\ddot{x}_i^*} + \underbrace{(\varepsilon_i - \beta_2 v_i)}_{\varepsilon_i^*}\end{aligned}$$

- Messfehler v sowohl im Regressor \ddot{x}^* als auch im Störterm ε^*
 $\Rightarrow E(\varepsilon^* | \ddot{x}^*) \neq 0 \Rightarrow \text{Endogenität!}$

Messfehler in den erklärenden Variablen

OLS Schätzung:

$$\begin{aligned}\hat{\beta}_2 &= \frac{\sum \ddot{x}^* \ddot{y}}{\sum (\ddot{x}^*)^2} \\ &= \frac{\sum \ddot{x}^* (\beta_2 \ddot{x} + \varepsilon)}{\sum (\ddot{x}^*)^2} \\ &= \beta_2 \frac{\sum (\ddot{x}^*) \ddot{x}}{\sum (\ddot{x}^*)^2} + \frac{\sum \ddot{x}^* \varepsilon}{\sum (\ddot{x}^*)^2}\end{aligned}$$

Messfehler in den erklärenden Variablen

OLS Schätzung:

$$\begin{aligned}\hat{\beta}_2 &= \frac{\sum \ddot{x}^* \ddot{y}}{\sum (\ddot{x}^*)^2} \\ &= \frac{\sum \ddot{x}^* (\beta_2 \ddot{x} + \varepsilon)}{\sum (\ddot{x}^*)^2} \\ &= \beta_2 \frac{\sum (\ddot{x}^*) \ddot{x}}{\sum (\ddot{x}^*)^2} + \frac{\sum \ddot{x}^* \varepsilon}{\sum (\ddot{x}^*)^2}\end{aligned}$$

$$\text{plim}(\hat{\beta}_2) = \text{plim} \left(\beta_2 \frac{\sum (\ddot{x}^*) \ddot{x}}{\sum (\ddot{x}^*)^2} + \frac{\sum \ddot{x}^* \varepsilon}{\sum (\ddot{x}^*)^2} \right) = \beta_2 \left(\frac{\sigma_{\ddot{x}}^2}{\sigma_{\ddot{x}}^2 + \sigma_v^2} \right)$$

für $\ddot{x}_i^* = \ddot{x}_i + v_i$ mit $v_i \sim \text{i.i.d.}(0, \sigma_{v_i}^2)$

\Rightarrow *'attenuation bias'*

Messfehler in den erklärenden Variablen

OLS Schätzung:

$$\begin{aligned}\hat{\beta}_2 &= \frac{\sum \ddot{x}^* \ddot{y}}{\sum (\ddot{x}^*)^2} \\ &= \frac{\sum \ddot{x}^* (\beta_2 \ddot{x} + \varepsilon)}{\sum (\ddot{x}^*)^2} \\ &= \beta_2 \frac{\sum (\ddot{x}^*) \ddot{x}}{\sum (\ddot{x}^*)^2} + \frac{\sum \ddot{x}^* \varepsilon}{\sum (\ddot{x}^*)^2}\end{aligned}$$

$$\text{plim}(\hat{\beta}_2) = \text{plim} \left(\beta_2 \frac{\sum (\ddot{x}^*) \ddot{x}}{\sum (\ddot{x}^*)^2} + \frac{\sum \ddot{x}^* \varepsilon}{\sum (\ddot{x}^*)^2} \right) = \beta_2 \left(\frac{\sigma_{\ddot{x}}^2}{\sigma_{\ddot{x}}^2 + \sigma_v^2} \right)$$

für $\ddot{x}_i^* = \ddot{x}_i + v_i$ mit $v_i \sim \text{i.i.d.}(0, \sigma_{v_i}^2)$

\Rightarrow *'attenuation bias'*

Hinweis: Meßfehler in y führt zu keiner Endogenität!

Ökonometrische Endogenität

Gauss-Markov Annahme $E(\varepsilon_i|\mathbf{X}) = 0$ verletzt!

- Mögliche Ursachen:

Ökonometrische Endogenität

Gauss-Markov Annahme $E(\varepsilon_i|\mathbf{X}) = 0$ verletzt!

- Mögliche Ursachen:
 - Fehlende relevante Variablen (*omitted variables*):
mit Sonderfällen

Ökonometrische Endogenität

Gauss-Markov Annahme $E(\varepsilon_i|\mathbf{X}) = 0$ verletzt!

- Mögliche Ursachen:
 - Fehlende relevante Variablen (*omitted variables*):
mit Sonderfällen
 - Selektionsprobleme

Ökonometrische Endogenität

Gauss-Markov Annahme $E(\varepsilon_i|\mathbf{X}) = 0$ verletzt!

- Mögliche Ursachen:
 - Fehlende relevante Variablen (*omitted variables*):
mit Sonderfällen
 - Selektionsprobleme
 - '*unobserved heterogeneity*'

Ökonometrische Endogenität

Gauss-Markov Annahme $E(\varepsilon_i|\mathbf{X}) = 0$ verletzt!

- Mögliche Ursachen:
 - Fehlende relevante Variablen (*omitted variables*):
mit Sonderfällen
 - Selektionsprobleme
 - '*unobserved heterogeneity*'
 - Simultane Kausalität (z.B. '*reverse causality*')

Ökonometrische Endogenität

Gauss-Markov Annahme $E(\varepsilon_i|\mathbf{X}) = 0$ verletzt!

- Mögliche Ursachen:
 - Fehlende relevante Variablen (*omitted variables*):
mit Sonderfällen
 - Selektionsprobleme
 - *'unobserved heterogeneity'*
 - Simultane Kausalität (z.B. *'reverse causality'*)
 - Messfehler in den erklärenden Variablen

Ökonometrische Endogenität

Gauss-Markov Annahme $E(\varepsilon_i|\mathbf{X}) = 0$ verletzt!

- Mögliche Ursachen:
 - Fehlende relevante Variablen (*omitted variables*): mit Sonderfällen
 - Selektionsprobleme
 - '*unobserved heterogeneity*'
 - Simultane Kausalität (z.B. '*reverse causality*')
 - Messfehler in den erklärenden Variablen
- Mögliche Lösungen

Ökonometrische Endogenität

Gauss-Markov Annahme $E(\varepsilon_i|\mathbf{X}) = 0$ verletzt!

- Mögliche Ursachen:
 - Fehlende relevante Variablen (*omitted variables*): mit Sonderfällen
 - Selektionsprobleme
 - '*unobserved heterogeneity*'
 - Simultane Kausalität (z.B. '*reverse causality*')
 - Messfehler in den erklärenden Variablen
- Mögliche Lösungen
 - Experimente mit Randomisierung

Ökonometrische Endogenität

Gauss-Markov Annahme $E(\varepsilon_i|\mathbf{X}) = 0$ verletzt!

- Mögliche Ursachen:
 - Fehlende relevante Variablen (*omitted variables*): mit Sonderfällen
 - Selektionsprobleme
 - *'unobserved heterogeneity'*
 - Simultane Kausalität (z.B. *'reverse causality'*)
 - Messfehler in den erklärenden Variablen
- Mögliche Lösungen
 - Experimente mit Randomisierung
 - Instrumentvariablen

Instrumentvariablen

Beispiel: Lernzeit und Lernerfolg

Zentrale Idee: Zusätzliche Information!

- z.B. Wetter am Vorabend W (oder Fussball Weltmeisterschaft) wirkt *ausschließlich* auf Lernzeit, und nur indirekt über die Lernzeit auf den Lernerfolg
- Erlaubt *zweistufige* Vorgangsweise

1

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

2

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

- *Methode der zweistufigen kleinsten Quadrate (2SLS)*
Erlaubt *konsistente* Schätzung von $\widehat{\beta}_2$!
- Intuition: Instrumentvariablen wirken wie *natürliche Experimente*!

Instrumentvariablen

‘Instrumentvariablen’ (z) müssen zwei Anforderungen erfüllen:

- ➊ **Relevanz:** $\text{cov}(z, x) \neq 0$
(sie müssen eine Wirkung haben)
- ➋ **Exogenität:** $\text{cov}(z, \varepsilon) = 0$ (*‘exclusion restriction’*)
(sie müssen wie eine *Intervention* in einem Experiment wirken, d.h., sie dürfen *ausschließlich* indirekt über x auf y wirken!)

Instrumentvariablen

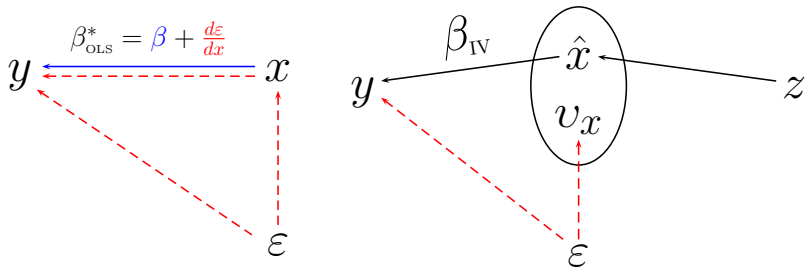
‘**Instrumentvariablen**’ (z) müssen zwei Anforderungen erfüllen:

① **Relevanz:** $\text{cov}(z, x) \neq 0$

(sie müssen eine Wirkung haben)

② **Exogenität:** $\text{cov}(z, \varepsilon) = 0$ (*‘exclusion restriction’*)

(sie müssen wie eine *Intervention* in einem Experiment wirken, d.h., sie dürfen *ausschließlich* indirekt über x auf y wirken!)



Instrumentvariablen

- Praktisch erfolgt Schätzung einstufig als
Methode der Momente Schätzer
- Einfacher Fall: Moment-Bedingungen für Grundgesamtheit:
 - Erwartungswert der Störterme ist Null

$$E(\varepsilon_i) = 0$$

- die Kovarianz zwischen den Störtermen ε_i und (stochastischen) x_i ist Null

$$E(x_i \varepsilon_i) = 0$$

- Die entsprechenden Bedingungen *für die Stichprobe*:

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

- Einsetzen der Stichprobenresiduen $\hat{\varepsilon}_i = y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i$ und Lösen der beiden Gleichungen führt wieder zur OLS Schätzfunktion.

Instrumentvariablen

IV als Methode der Momente Schätzer

- **IV-Schätzfunktion:** Momentbedingungen

$$\frac{1}{n} \sum \varepsilon_i = 0 \quad \text{und} \quad \frac{1}{n} \sum (z_i \varepsilon_i) = 0$$

Einsetzen der IV-Residuen $\hat{\varepsilon}_i = y_i - \hat{\beta}_1^{\text{IV}} - \hat{\beta}_2^{\text{IV}} x_i$ in die Stichproben-Momentbedingungen

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

und lösen gibt

$$\hat{\beta}_2^{\text{IV}} = \frac{\sum (y_i - \bar{y})(z_i - \bar{z})}{\sum (x_i - \bar{x})(z_i - \bar{z})} = \frac{\text{cov}(y, z)}{\text{cov}(x, z)}$$

$$\hat{\beta}_1^{\text{IV}} = \bar{y} - \hat{\beta}_2^{\text{IV}} \bar{x}$$

Instrumentvariablen

IV als Methode der Momente Schätzer

$$\hat{\beta}_2^{\text{IV}} = \frac{\text{COV}(y, z)}{\text{COV}(x, z)}$$

$$\hat{\beta}_1^{\text{IV}} = \bar{y} - \hat{\beta}_2^{\text{IV}} \bar{x}$$

Probleme:

- Direkter Test auf Gültigkeit der Instrumente (*exclusion restriction*) nicht möglich!
- Schwache Instrumente ($\text{cov}(x, z)$ klein)
- IV Schätzer sind meist 'ungenau' (große Varianz)

Hausman Test

- Liegt Endogenität vor? $H_0 : \text{plim} \frac{1}{n} \sum_i \ddot{x}_i \varepsilon_i = 0$
(ist IV Schätzung erforderlich?)
- Vergleich zweier Schätzfunktionen:
 - ➊ sowohl unter der Null-Hypothese als auch unter der Alternativhypothese konsistent (z.B. IV - Schätzfunktion)
 - ➋ nur unter Nullhypothese konsistent
- Ein großer Unterschied wird als Evidenz zugunsten der Alternativhypothese interpretiert

$$H = \frac{\left(\hat{\beta}_h^{\text{IV}} - \hat{\beta}_h^{\text{OLS}} \right)^2}{\widehat{\text{var}} \left(\hat{\beta}_h^{\text{IV}} \right) - \widehat{\text{var}} \left(\hat{\beta}_h^{\text{OLS}} \right)} \stackrel{a}{\sim} \chi_1^2$$

- Setzt gültige Instrumente für Schätzung von $\hat{\beta}_h^{\text{IV}}$ voraus!!!

Test auf überidentifizierende Restriktionen

- Direkter Test der *exclusion restriction* ($\text{cov}(\varepsilon, z) = 0$, d.h., IV wirken *ausschließlich* über x auf y) nicht möglich!
- Falls
 - ➊ mehr IV zur Verfügung stehen als erforderlich (überidentifiziert), und
 - ➋ mindestens so viele IV *gültig* sind wie für eine konsistente Schätzung benötigt werdenkönnen die überidentifizierenden IV getestet werden
- z.B. *Sargan* Test, *J*-Test, ...

Praktische Hinweise

- Gute Instrumente sollten etwas ähnliches leisten, wie die Intervention in einem Experiment, sie dürfen *nur* über die Regressoren x auf y wirken!

Praktische Hinweise

- Gute Instrumente sollten etwas ähnliches leisten, wie die Intervention in einem Experiment, sie dürfen *nur* über die Regressoren x auf y wirken!
- Woher kommen 'gute' Instrumente?

Praktische Hinweise

- Gute Instrumente sollten etwas ähnliches leisten, wie die Intervention in einem Experiment, sie dürfen *nur* über die Regressoren x auf y wirken!
- Woher kommen 'gute' Instrumente?
 - Gesetzliche Regelungen, Regulierung, etc. Diese sollten das 'Assignment' unabhängig vom 'Outcome' beeinflussen.

Praktische Hinweise

- Gute Instrumente sollten etwas ähnliches leisten, wie die Intervention in einem Experiment, sie dürfen *nur* über die Regressoren x auf y wirken!
- Woher kommen 'gute' Instrumente?
 - Gesetzliche Regelungen, Regulierung, etc. Diese sollten das 'Assignment' unabhängig vom 'Outcome' beeinflussen.
 - Natur (Geographie, Klima (Wetter), Biologie, ...)

Praktische Hinweise

- Gute Instrumente sollten etwas ähnliches leisten, wie die Intervention in einem Experiment, sie dürfen *nur* über die Regressoren x auf y wirken!
- Woher kommen 'gute' Instrumente?
 - Gesetzliche Regelungen, Regulierung, etc. Diese sollten das 'Assignment' unabhängig vom 'Outcome' beeinflussen.
 - Natur (Geographie, Klima (Wetter), Biologie, ...)
 - Falls verfügbar, RCT (*Randomized Controlled Trials*; → 'intention to treat')

Praktische Hinweise

- Gute Instrumente sollten etwas ähnliches leisten, wie die Intervention in einem Experiment, sie dürfen *nur* über die Regressoren x auf y wirken!
- Woher kommen 'gute' Instrumente?
 - Gesetzliche Regelungen, Regulierung, etc. Diese sollten das 'Assignment' unabhängig vom 'Outcome' beeinflussen.
 - Natur (Geographie, Klima (Wetter), Biologie, ...)
 - Falls verfügbar, RCT (*Randomized Controlled Trials*; → '*intention to treat*')
 - Opportunitätskosten, die sich zwischen Individuen unterscheiden;

Praktische Hinweise

- Gute Instrumente sollten etwas ähnliches leisten, wie die Intervention in einem Experiment, sie dürfen *nur* über die Regressoren x auf y wirken!
- Woher kommen 'gute' Instrumente?
 - Gesetzliche Regelungen, Regulierung, etc. Diese sollten das 'Assignment' unabhängig vom 'Outcome' beeinflussen.
 - Natur (Geographie, Klima (Wetter), Biologie, ...)
 - Falls verfügbar, RCT (*Randomized Controlled Trials*; → '*intention to treat*')
 - Opportunitätskosten, die sich zwischen Individuen unterscheiden;
 - ...

Praktische Hinweise

- Gute Instrumente sollten etwas ähnliches leisten, wie die Intervention in einem Experiment, sie dürfen *nur* über die Regressoren x auf y wirken!
- Woher kommen 'gute' Instrumente?
 - Gesetzliche Regelungen, Regulierung, etc. Diese sollten das 'Assignment' unabhängig vom 'Outcome' beeinflussen.
 - Natur (Geographie, Klima (Wetter), Biologie, ...)
 - Falls verfügbar, RCT (*Randomized Controlled Trials*; → 'intention to treat')
 - Opportunitätskosten, die sich zwischen Individuen unterscheiden;
 - ...
- Für die Wahl 'guter' Instrumente ist meist eine fundierte Kenntnis des datengenerierenden Prozesses (DGP) erforderlich!

Praktische Hinweise

Interne Validität

Die '*exclusion restriction*' kann nicht direkt empirisch getestet werden, deshalb muss '*Überzeugungsarbeit*' geleistet werden!

Wie? Gründliches Nachdenken (ökonomische Theorie)!

- Stellen Sie sich ein (randomisiertes) Experiment vor, welches geeignet wäre, die Frage zu beantworten;
- Etablierte Normen und Konventionen, denen in gut publizierten (!) Studien gefolgt wird;

Praktische Hinweise

Interne Validität

Sowohl für die Spezifikation als auch für die Publikation beachten Sie folgende Punkte:

- Diskutieren Sie die identifizierenden Annahmen!
- Überprüfen und diskutieren Sie die *reduzierte Form*: wenn Sie dort nichts finden, ist möglicherweise nichts da.
- Überprüfen und diskutieren Sie die *1. Stufe*: stimmen die Vorzeichen und Größenordnungen mit den Erwartungen überein?
- Die F-Statistik der 'ausgeschlossenen' (*excluded*) Instrumente sollte mindestens größer als 10 sein
(→ schwache Instrumente)
- Überprüfen und diskutieren Sie den Unterschied zwischen OLS und IV Schätzung. Entspricht der Unterschied den Erwartungen? Kein HARKing!
- Diskutieren Sie auch externe Validität.

Kausalität II

- **Kontrafaktische Situation:** was wäre passiert, wenn das verursachende Ereignis *nicht* eingetreten wäre?
- Aber: *Faktum* und *Kontrafaktum* können nie am gleichen Individuum beobachtet werden!
- Entscheidende Frage: was ist die Referenzgruppe? (z.B. *treatment* und *control* Gruppe)
- Hypothesentest per se erlauben keine Rückschlüsse auf Kausalität, dazu bedarf es in der Regel des “*design of an experiment*” → **Identifikation**

Kausalität II

- **OLS:** Koeffizienten könne kausal interpretiert werden, wenn die PRF korrekt spezifiziert wurde und den DGP vollständig und korrekt beschreibt!
Das impliziert u.a.:
 - keine fehlenden relevanten Variablen (d.h. auch keine Selektionsprobleme und keine unbeobachtete Heterogenität)
 - keine unberücksichtigten Interdependenzen (d.h. keine simultane Kausalität, keine *omitted equations*)
 - keine Meßfehler bei Regressoren
- Dies kann de facto nie überprüft werden!
- Deshalb sollten OLS Regressionen nicht kausal interpretiert werden!!!
- **Hinweis:** Alle erwähnten Tatbestände führen zu einer Korrelation der Störterme mit den Regressoren (d.h. $E(\varepsilon_i|\mathbf{X}) \neq 0$)

Kausalität II

Wann sollten die Alarmglocken läuten?

- wenn eine oder mehrere nicht berücksichtigte Variablen vorstellbar sind, die sowohl mit y als auch einer berücksichtigten x Variable korreliert sein könnten → '*omitted variables*'
z.B. Einkommen – Bildung | Fähigkeiten
- wenn eine x Variable das Ergebnis einer Entscheidung ist → Simultane Kausalität!
z.B. *gender-wage gap* und Branche
- **Achtung:** ein einziger endogener Regressor verdirbt den Brei (d.h. führt dazu, dass *alle* Koeffizienten verzerrt sind)!

Kausalität: mögliche Maßnahmen

Instrumentvariablen:

- Ausnützung einer *exogenen* Variation, die einer pseudo-experimentellen Intervention entspricht.
- *Exogenität* kann nicht direkt empirisch überprüft werden!
- Geringe Verletzung der Annahmen kann dramatische Folgen haben!
- kein Allheilmittel . . .

Kausalität: mögliche Maßnahmen

Fixed Effects Modelle:

- wiederholte Beobachtungen für Individuen (Panel) erlauben individualspezifische Dummies (*'fixed effects'* und *'Least Squares dummy Variable'* Modell sind numerisch äquivalent, siehe FWL Theorem)
- Dummy Variablen kontrollieren für alles, was sich über die Zeit nicht verändert!
- Mehr Kontrolle, aber löst grundsätzliches Problem nicht!

Kausalität: mögliche Maßnahmen

Difference-in-difference Modelle:

- doppelter Vergleich: *vorher* – *nachher* und *treatment* – *control* Gruppe.
- Falls Kontrollgruppe tatsächlich das '*counterfactual*' abbildet sind Kausalaussagen möglich.
- Aber es kann kaum überprüft werden, inwieweit die Kontrollgruppe Kontrollgruppe tatsächlich das '*counterfactual*' das '*counterfactual*' beschreibt.

Kausalität: mögliche Maßnahmen

Regression Discontinuity Modelle:

- Ausnützung einer exogenen Schwelle (z.B. Teilungsziffern in Schulklassen) und Vergleich der knapp darunter und knapp darüber liegenden Gruppe.
- Spezifikations- und Selektionsprobleme?
- Begrenzt einsetzbar
- auch dies kein Allheilmittel ...



So Long, and Thanks for All the Fish ...

Grundlagen der Ökonometrie

herbert.stocker@uibk.ac.at

www.hsto.info/econometrics