emp WiFo

Aufg 60

„z sei ein geeignetes Instrument", d.h. z erfüllt insbesondere die folgenden Annahmen:

1. Annahme $Cov(z, \epsilon) = 0$

↳ unkorreliert, z ist exogene Var.

↳ Annahme muss durch theoretisch begründet werden, kann nicht getestet werden (wegen ϵ)

2. Annahme $Cov(z, X) \neq 0$

↳ korreliert und möglichst hoch, da wir mit z einen Teil von X erklären wollen

↳ Annahme kann getestet werden mittels OLS-Regression (Einfach-Regression)

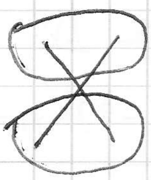
Einfach-Regression:

Situation wahr:

$$\begin{array}{ccc} X & \longrightarrow & Y \\ \downarrow & & \\ \epsilon & & \end{array}$$

Problem: korreliert

Idee:



→ gesunder Teil, der mit z korreliert, aber nicht mit ϵ wird genutzt

→ „fauler" Teil, der mit ϵ korreliert ist, wird abgeschnitten

Situation nachher:

hoch korreliert

$$\begin{array}{ccccc} X & \longleftrightarrow & Z & \longrightarrow & Y \\ & \searrow & \vdots & & \\ & & \epsilon & & \end{array}$$

nicht korreliert

Problem wird unwichtig

· formale Umsetzung durch 2malige OLS-Anwendung:

Ausgangsmodell: $Y = \beta_0 + \beta_1 X + \epsilon$ z sei Instrument für X

1. OLS-Regression: $x = \pi_0 + \pi_1 \cdot z + v$

Schätzung ergibt $x^* = \hat{\pi}_0 + \hat{\pi}_1 \cdot z$

außerdem Signifikanztest von z bzw. π_1 mittels t-Test überprüfen,

d. h. teste $H_0: \pi_1 = 0$

π_1 muss signifikant von Null verschieden sein, damit Annahme 2 erfüllt ist.

2. OLS-Regression $y = \beta_0 + \beta_1 x^* + \varepsilon$

nochmals schätzen $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x^*$

diese Gleichung ist um den endogenen Teil von x bereinigt, es ist nur noch der „gesunde" Teil enthalten.

Warum $H_0: \pi_1 = 0$ testen. Erst wenn $\pi_1 \neq 0$ ist, gilt Annahme 2.

Def: $\pi_1 = \dfrac{Cov(z,x)}{Var\, z}$

$$Cov(z,y) = Cov(z, \beta_0 + \beta_1 x + \varepsilon)$$

$$= \underbrace{\beta_1 cov(z,x)}_{\neq 0 \text{ nach Annahme 2}} + \underbrace{cov(z,\varepsilon)}_{=0 \text{ nach Annahme 1}}$$

$\Leftrightarrow \beta_1 = \dfrac{Cov(z,y)}{Cov(z,x)}$ um dieses β_1 zu schätzen benutzen wir analog den IV-Schätzer für β_1:

$$\hat{\beta}_{1_{IV}} = \frac{\widehat{cov}(z,y)}{\widehat{cov}(z,x)} = \frac{\frac{1}{n-1}\sum(z_i - \bar{z})(y_i - \bar{y})}{\frac{1}{n-1}\sum(z_i - \bar{z})(x_i - \bar{x})}$$

$$\hat{\beta}_{0_{IV}} = \bar{y} - \hat{\beta}_1 \bar{x}$$

Man erkennt: Die Formel für $\hat{\beta}_{1_{IV}}$ (also IV-Schätzer) ergibt mit $z = x$ den OLS-Schätzer für $\hat{\beta}_1$.

$$\hat{\beta}_1 = \frac{\frac{1}{n-1}\sum(x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n-1}\sum(x_i - \bar{x})^2} = \frac{Cov(x,y)}{Var(x)} = \hat{\beta}_{1_{OLS}}$$

Um Konsistenz zu zeigen, betrachten wir wieder das W'keitslimit:

$$p\lim(\hat{\beta}_1) = p\lim \frac{\widehat{Cov}(z,y)}{\widehat{Cov}(z,x)}$$

$$= \frac{\text{p lim } \widehat{cov}(z,y)}{\text{p lim } \widehat{cov}(z,x)}$$

$$= \frac{\text{p lim } \frac{1}{n-1} \sum (z_i - \bar{z})(y_i - \bar{y})}{\text{p lim } \frac{1}{n-1} \sum (z_i - \bar{z})(x_i - \bar{x})}$$

| vgl. Ü 7

Satz der großen Zahlen:

$\bar{z} \rightarrow E[z] \quad \Rightarrow \quad = \frac{Cov(z,y)}{Cov(z,x)} = \beta_1$

Aufg 61

a) Begründung für Korrelation zwischen Kurs und ε:

Es ist bekannt, dass der sozioökonomische Hintergrund einen Einfluss auf die Studienleistung hat.

In ε kann z.B. das Familieneinkommen, Bildungsgrad der Eltern enthalten sein, welches ein positiven Effekt auf Punkte hat und wahrscheinlich mit der Kursteilnahme korreliert ist.

b) ~~Kurs~~ Einkommen als IV-Schätzer geeignet?

↳ Annahme 2: Cov(Einkommen, Kurs) $\neq 0$ ✓

↳ Annahme 1: Cov(Einkommen, ε) $= 0$ ↯ (siehe a)

$\Rightarrow$ ~~Kurs~~ Einkommen nicht als IV-Schätzer geeignet.

c) Definiere Dummy-Variable Stipendium, mit Stipendium $=1$, dann Gebühren erlassen, sonst 0

$\rightarrow$ Wegen rein zufälliger Vergabe der Stipendien, muss gelten, Cov(Stipendium, ε) $= 0$

$\rightarrow$ Außerdem sollte Cov(Stipendium, Kurs) $\neq 0$ gelten

$\Rightarrow$ zufällige Vergabe von Stipendien ein IV für Kurs

Aufg 62

a) kontrolliere: Familieneinkommen, Bildungsabschluss der Eltern...

b) Modell: Punkte $= \beta_0 + \beta_1 \cdot$ Lyzeum $+ \beta_2 \cdot$ Familieneinkommen $+ \beta_3 \cdot$ Vater Ausb. $+ \beta_4 \cdot$ Mutter Ausb $+ \varepsilon$

c) Cor(E, Lyzeum) = 0 ?

↳ theoretische Begründung: Eltern, die hoch motiviert sind und ihre Töchter beim Schulabschluss unterstützen wollen, könnten eher dazu geneigt sein ihre Töchter an einem Mädchen-gymnasium anzumelden

⇒ Cor zwischen E und Lyzeum wahrscheinlich!

d) AnzMG sei die Anzahl der Mädchengymnasien im Umkreis von 30 km.

Wenn AnzMG eine gültige IV für Lyzeum sein soll, müssen die beiden Annahmen erfüllt sein:

1) Cor(AnzMG, E) = 0

2) Cor(AnzMG, Lyzeum) ≠ 0

zu 2) Mit steigender Anzahl von Mädchengymnasien wird es für Mädchen wahrscheinlicher ein solches zu besuchen

⇒ Ann 2 ✓

zu 1) problematischer

Einige Regionen können im Durchschnitt bessere Schülerinnen aus Gründen haben, die nicht mit dem Familieneinkommen / Ausbildung der Eltern zu tun haben.

⇒ theoretische Begründung von Annahme 1 anzuzweifeln

<u>Aufg 63</u>

a) Begründe Cor(PC, E) ≠ 0:

↳ in E wird wahrscheinlich das Familieneinkommen enthalten sein

b) Einkommen als Instrument für PC geeignet?

2. Annahme ✓ (höheres Einkommen macht Besitz wahrscheinlicher)

1. Annahme Cor(Einkommen, E) = 0, ist wahrscheinlich nicht erfüllt, da das höhe Einkommen mit anderen sozioökonomische Faktoren, die weiterhin ~~im~~ E sind,

korreliert ist ↯

⇒ Einkommen ist nicht als IV für PC geeignet

c) Gutschein als IV für PC geeignet?

1. Annahme: Cor(Gutschein, ε) = 0, ist erfüllt, da der Gutschein rein zufällig vergeben wurde

2. Annahme: Cor(Gutschein, PC) ≠ 0, ist testbar:
Studenten, die einen Gutschein für einen geschenkten PC bekommen, werden diesen annehmen
→ 2. Ann erfüllt

⇒ Gutschein ist IV für PC

Aufg 64

sibs ≙ siblings = Anzahl der Geschwister

a) OLS-Regression von ln(wage) auf sibs ergibt mittels R-Project:

$$\widehat{\ln(wage)} = \underset{(0{,}022)}{6{,}861} - \underset{(0{,}0059)}{0{,}0279} \cdot sibs$$

$n = 935 \quad R^2 = 0{,}023$

Wenn ein Geschwisterkind c.p. hinzu kommt, verringert sich der Lohn um ca. 2,5%.

Die t-Statistik ist $\frac{-0{,}0279}{0{,}0059} = -4{,}728$ und somit hoch signifikant

⇒ β_1 ist von Null verschieden

IV-Regression von ln(wage) auf educ, wobei wir sibs als Instrument für educ benutzen:

$$\widehat{\ln(wage)} = \underset{(0{,}355)}{5{,}130} + \underset{(0{,}026)}{0{,}122} \cdot educ$$

$n = 935 \quad R^2 < 0$

- Dies zeigt, dass c.p. ein weiteres Jahr Bildg. zu einem höheren Gehalt von 12,2% führt.

b) Begründe $Cor(educ, brthord) < 0$:

Es könnte sein, dass ältere Kinder bevorzugt Zugang zu höherer Bildung bekommen und Familien aufgrund beschränkter Mittel später geborene Kinder weniger Bildg. finanzieren können.

Einfach Regression von educ auf brthord mit R

$$\widehat{educ} = \underset{(0{,}13)}{14{,}15} - \underset{(0{,}046)}{0{,}283} \cdot brthord$$

Man erkennt, dass bei Erhöhg. von brthord um eine Einheit, die Ausbildungszeit um ca. 0,233 Jahre sinkt.