91,5 p A

Fråga 1 Anonymkod: 0051-SYM Blad 1

1) 2SLS/IV regression: C 3

2) normally distributed: A 3

3): A 3

4): A 3

5): A 3

6): C 3

tot: 18/18

Fråga (2) Anonymkod: 0051-SYM Blad (2)

Vi har följande regression:

$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{1i}^2 + \beta_3 \log x_{2i} + u_i \quad u_i \sim N(0, \sigma_i^2)$

Vi har också noterat att våran slump variabels varians är beroende av $\log x_{i2} \Rightarrow \sigma_{u_i}^2 = \sigma^2 \times \log x_{i2}$

A) Jag kommer ha samma hypoteser för alla koefficienter.

$H_0: \beta_j = 0 \quad H_a: \beta_j \neq 0 \quad \alpha = 0{,}01$

Detta kommer göra så att teststatistikan blir följande $\frac{\beta_j}{se(\beta_j)}$ under H_0 ok

och jag kommer förkasta H_0 för respektive koefficient om $|t_{obs_j}| > |t_{\alpha/2}(n-k-1)|$ vårat kritiska värde är samma för alla och det är $t_{0{,}005}(296) = 2{,}576$ ok

β_0: $\frac{5}{0{,}3} = 16{,}667$ β_0 är statistisk signifikant skilt från 0

β_1: $\frac{1{,}4}{0{,}56} = 2{,}5$ β_1 är inte statistisk signifikant skilt från 0

β_2: $\frac{-0{,}125}{0{,}1} = -1{,}25$ β_2 är inte skilt från 0

4

β_3: $\frac{6{,}2}{1{,}1} = 5{,}6363$ β_3 är skilt från 0.

B) $\frac{\Delta Y}{\Delta x_1} = \beta_1 - 2\beta_2 x_{1i}$

Detta betyder att om vi ändrar x_1 med en enhet så ändras y beroende på vart vi redan står i x_1 !!

$x^* = \left|\frac{\beta_1}{2\beta_2}\right| = 5{,}6 = \left|\frac{1{,}4}{2 \cdot 0{,}125}\right|$

4

x^* är vändningspunkten. Om vi ökar x_1 efter det kommer Y påverkas negativt.

(Vänd Blad)

Fråga 2 Anonymkod: 0051-SYM Blad: 3

c) eftersom x_2 är i log form kommer det se "annorlunda" ut.

$\Delta \hat{y} = (\beta_3/100)\% \Delta x_2$ Detta betyder att om

x_2 ökar med 1% så kommer y att öka

med $\frac{6,2}{100} = 0,062$ Frågas efter en enhets förändring i x_2... 3

not: Man skalle möjligen kunna anta att Δy givet Δx_2 också beror på u_i då $\sigma^2_{u_i} = \sigma^2 \times \log x_{2i}$

D) Vi vill prediktera $\hat{y}_i$ och göra ett prediktions-intervall. Givet $x^0_{1i} = 4$ $x^0_{2i} = 30$

Jag kallar det skattade värdet av y för $\hat{y}^0$

där $\hat{y}^0 = \beta_0 + \beta_1 x^0_{1i} + \beta_2 x^{2\,0}_{2i} + \beta_3 \log x^0_{2i}$ med våra estimerade koefficienter blir det följande:

$\hat{y}^0 = 5 + 1,4 \cdot 4 - 0,125 \cdot 4^2 + 6,2 \log(30) = 7 - 2 + 21,087 = 26,087$

5,6 rätt tänkt, fel uträknat.

$\hat{y}^0 = 26,087$

standardavvikelsen för $\hat{y}^0$ är $\sqrt{[se(\hat{y}^0)]^2 + \hat{\sigma}^2}$

$\hat{\sigma}^2 = 0,25$

men $se(\hat{y}^0)$ får man genom att köra regressionen

y på $(x_{1i} - c_1), (x_{2i} - c_2) \dots (x_{ki} - c_k)$ när man vill

prediktera medelvärdet av y. Då får man $se(\hat{y}^0)$ från

interceptet, vilket jag kommer använda nu. $se(\hat{\beta}) = 0,3$

$\sqrt{0,3^2 + 0,25} = \sqrt{0,34} = 0,583$

eftersom vi ska ha ett 95% intervall kommer jag

använda 1,96 (från tabell)

$26,087 \pm 1,96 \cdot 0,583 = 26,087 \pm 1,14268$ bra förklarat trots min miss!

$(24,94432 ; 27,5138)$ 3

Fråga (2) Anonymkod: 0051-SYM Blad: (4)

E) $\tilde{y} = 2y + 3 = 2(\beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 \log x_2 + u) + 3$

eftersom att 3 är en konstant så kommer den att ändra på interceptet $\tilde{\beta}_0 = \beta_0 - 3$ blir +

och eftersom att y är multiplicerat med 2 kommer alla β_j också att multipliceras med 2

Förutom β_3 detta eftersom att koefficienter i log berörs inte av "omskalningar" (detta gäller omskalningar i x-variabeln.

Varken t eller F värdena kommer att påverkas av detta. ok, bra

(2)

F) 1 = Rätt 2

2 = Fel 2

3 = Fel 2

(6)

G) För att lösa problemet med heteroskedasititet kommer jag använda WLS. Detta eftersom jag vet här $h(x)$ "är att" $(\log x_{2i})$. Det man gör är att multiplicera alla termer med $\frac{1}{\sqrt{\log(x_{2i})}}$ även y. Detta leder till att vi kommer vikta ner de obs som har stor varians. Formellt ser det ut följande: $\min \sum (y_i - \hat{y}_i)^2 / \log x_{2i}$

$u^* = \frac{u}{\sqrt{\log x_{2i}}}$

$Var(u|x) = E(u^2|x) - [E(u|x)]^2 = E(u^2|x)$

$E(u^{*2}|x) = E\left[\left(\frac{u_i}{\sqrt{\log x_{2i}}}\right)^2 \middle| x_i\right] = \frac{E(u_i^2|x_i)}{\log x_{2i}} = \frac{\sigma^2 \log x_{2i}}{\log x_{2i}} = \sigma^2$

(3)

fördel: högre effektivitet

tot: 25/30

Fråga (3)

Anonymkod: 0051-SYM Blad (5)

A) Det står inte har "wage" nått men antar att det är dollar/timme

Vi kan se att det är ett positivt samband flyr år med samma arbetsgivare ger en högre lön. Den är signifikan vilket är bra. om man skulle jobba ett år till hos sin arbetsgivare skulle ens lön öka med 0,017 · 100% = 1,7% dollar/timmen. det är en liten ökning men R^2 är 0,037 vilket tyder på att tenure inte förklarar lönen så bra (själv). Sen värderar nog arbetsgivaren "hur bra man utför jobbet" mer än lojalitet.

(4)

B) om man jämför en som jobbat 4 år respektive 7 blir det följande.
den som har jobbat i 3 mer år $\Delta y = (\beta_1 \cdot 100) \cdot \Delta x_1 \Rightarrow (0,013 \cdot 100) \cdot 3 = 1,3 \cdot 3 = 3,9\%$. Den som jobbat i 7 år tjänar 3,9% mer i lön.

(5)

C) Ja! det är bra. I verkligheten är det sällan så enkelt att förklara något med endast en variabel. Det blir mer realistiskt med fler variabler. Genom att inkludera fler variabler kan man lösa problem med bias. genom att inkludera variabler i modellen "försvinner" de från u.
Det som däremot skulle vara dåligt är att variansen skulle öka med fler variabler (allt lika). och risk för multikolinaritet finns.
om man vet att det inte finns multikolinaritet så är lite högre varians att föredra än bias.

(4)

Fråga 3

Anonym Kod: 0051-SYM

Blad: 6

D) Vi har modellen: $\log(wage) = \beta_0 + \beta_1 tenure + \beta_2 educ + \beta_3 IQ + \beta_4 exper + \beta_5 exper^2 + \beta_6 age + u$.

och vi vill testa om $\beta_3, \beta_4, \beta_5$ är signifikanta tillsammans. Detta gör vi genom ett F-test.

$H_0: \beta_4 = \beta_5 = \beta_6 = 0$ H_a: H_0 är inte sann. $\alpha = 0{,}05$

$$F = \frac{(R^2_{ur} - R^2_r)/q}{(1-R^2_{ur})/(n-k-1)}$$

ok

$R^2_r = 0{,}092$ $R^2_{ur} = 0{,}097$ $q = 3$ $df = 300-6-1 = 293$

Kritiskt värde $F_{3,293}(0{,}05) = F_{3,\infty}(0{,}05) = 2{,}6$

0,43640 ?

→ något räknefel

Om $F_{obs} > 2{,}6$ så förkastas H_0

$$F = \frac{(0{,}097-0{,}092)/3}{(1-0{,}097)/293} = \frac{0{,}00166667}{0{,}00381913} = 0{,}43640.$$

Vi kommer inte att förkasta H_0. Detta betyder att β_4 β_5 β_6 inte är signifikanta tillsammans.

4,5

Detta är logiskt om man kollar på R^2_{adj} som är $R^2_{adj}(ur) = 0{,}078$ $R^2_{adj}(r) = 0{,}083$.

Den mindre modellen är "bättre".

E) $t = \frac{\beta_j}{se(\beta_j)} = \frac{0{,}198}{0{,}051} = 3{,}88235$ $H_0: \beta_j = 0$ $H_a: \beta_j \neq 0$

Vi ser för olika signifikans grader.

3

$\alpha = 0{,}05$ $df = 300-10-1 = 289$ $t_{0{,}025}(289) = 1{,}96$

$\alpha = 0{,}01$ $t_{0{,}005}(289) = 2{,}576$

Vi ser att urban är signifikant till och med på 1% nivån.

Dom som bor i ett urbant område tjänar 19,8% mer i snitt i lön.

$\%\Delta Y = 100 \times 0{,}198 = 19{,}8\%$

För en given nivå av educ, exper, IQ, etc.

Fråga ③ Anonymkod: 0051-SYM Blad ⑦

F) $H_0: \beta_{10} = 0{,}15$ $H_a: \beta_{10} > 0{,}15$ $\alpha = 0{,}05$

$\beta_{10} = 0{,}260$ $se(\beta_{10}) = 0{,}072$

$t = \dfrac{0{,}26 - 0{,}15}{0{,}072}$ givet H_0 $\Rightarrow \dfrac{0{,}11}{0{,}072} = 1{,}527778$

Vi förkastar för stora värden.

$t_{0{,}05}(289) = 1{,}64$

Vi kommer inte förkasta H_0 då $1{,}527778 < 1{,}64$

Detta betyder att vi kan inte utesluta att den verkliga effekten av β_{10} är 0,15 den genomsnittliga ölöningen för gifta skulle kunna vara 15 %.

5

tot: 25,5/27

Fråga (4) Anonymkod: 0051-SYM Blad: (8)

A) som det står i texten är variabeln "mkids" correlerad med slump. variabeln. $corr(x_1, u) \neq 0$. detta är inte bra då vi får bias estimatorer. Då använder man i detta fall två "instrument" variabler.
Z_1 = boyboy och Z_2 = Girl Girl. och för att de ska vara bra IV:s så ska $corr(z_1, u) = corr(z_2, u) = 0$ och $corr(z, x_1) \neq 0$ samt $corr(z_2, x_1) \neq 0$ man vill ha så hög corr som möjligt.
man är i sönerhet ute efter folk med mer än 2 barn och om man har fått 2 barn av samma kön så vill man nog ha ett till och hoppas det inte har samma kön som de andra.
och man kan också anta att eftersom "mkids" går efter antal barn medan våra IV:s går efter kön att de inte heller är correlerade med u. Intuitivt bra IV:s om man skulle anta att MLR 1-6 utom 4 är uppnått så kan man ändå estimera koefficienterna konsistenta.

(6)

B) JA! instrument relevans är att koefficienterna på våra IV:s när man kör regressionen "mkids" på alla exogena variabler ska vara signifikanta.
vi kan se i "reduced regression" följande

$0{,}079^{xxx}$	$0{,}059^{xxx}$
(0,003)	(0,003)
GirlGirl	boyboy

och eftersom att de har 3 stjärnor respektive är deras P-värde < 0,01 alltså är de signifikanta på 1% nivån. och faktiskt har relevans att förklara "mkids"

(4)

Fråga (4) Anonymkod: 0051-SYM Blad: (9)

c) JA! Vi kan testa det i detta fall då vi har en "över identifierad" model.
För att testa om $corr(z_j; x_i) = 0$ så ska vi köra två separata "reduced" regressioner, ex. om vi har en endogen variabel och 2 instrument.

$x_1 = \pi_0 + \pi_1 z_1 + r_1$ och $x_1 = \tilde{\pi}_0 + \tilde{\pi}_1 z_2 + r_2$

5

och i dessa två regressioner jämför vi sen vilken koefficient β för x_1 får. om det tillexempel har olika tecken eller bara skiljer sig åt mycket kan vi anta att någon av z_1 eller z_2 är correlerad med u. givet att alla andra antagandena är uppfylda.

eller båda

D) Jag skulle inte säga att de praktiskt skilldes. 2SLS: -0,119 och OLS: -0,129. skilnaden på dessa är endast 0,01. detta betyder alltså att om vi skulle använda 2SLS så skulle en kvinna ha 1% högre chans att vara med i arbetskraften. om det var runt 10% skillnad skulle jag nog reagera.

5

Om skillnaden är statistiskt signifikant olika alltså $H_0: \beta_{OLS} - \beta_{2SLS} = 0$ $H_a: \beta_{OLS} - \beta_{2SLS} \neq 0$
det finns ingen statistisk skillnad heller.

om vi tar hela USAs befolkning så är den ändå ganska många kvinnor det handlar om.

Fråga 4 Anonymkod.0051-SYM Blad: 10

E) När man kör 2SLS "reduced regression" så kör man den variabeln som är endugenes på alla exogena variabler plus instrument variablerna.
Vi vill ha effekten av x på y men tyvärr så är korr(x,u) ≠ 0.
om vi då kör $x = \pi_1 z_1 + \pi_2 z_2 \ldots\ldots + \pi_k z_k$
där alla z är exogena så kommer $\hat{x}$ att vara exogen och inhålla så mycket info om x som möjligt. När vi sen kör y på $\hat{x}$ så får vi den exogena effekten av x.

~~4~~ 3

+ att $\hat{\beta}_{2SLS}$ kan bli inkonsistent annars.

~~6~~t: 2~~3~~/25