

Econometrie: tussentijdse test 2

Oplossingen

Vraag 1

1. $SER = 0.516$: de relatieve variaties van $wage$ t.o.v. de verwachte waarde van $wage$, gegeven de waarde van $exper$ en $black$, zijn van de orde van meer dan 50% (of, naukeuriger, van de orde van $e^{0.516} - 1 = 0.68 = 68\%$). Opmerking: “log-dollar” is betekenisloos.
2. Het marginaal effect is 0.0641, d.w.z. dat voor iemand met reeds 5 jaar werkervaring een bijkomend jaar werkervaring een gemiddelde stijging het loon van 6.41% met zich meebrengt. De onderliggende berekening is als volgt:

$$\frac{\partial E(\log(wage)|black, exper)}{\partial exper} = \beta_2 + 2\beta_3 exper.$$

Met de schattingen $\hat{\beta}_2$ en $\hat{\beta}_3$ en $exper = 5$ is het geschat effect

$$\begin{aligned}\frac{\partial E(\log(wage)|black, exper)}{\partial exper} &= \hat{\beta}_2 + 2\hat{\beta}_3 \times 5 \\ &= 0.0641.\end{aligned}$$

3. De regressie in (1) is ongeschikt om deze vraag te beantwoorden. De reden is dat $\hat{\beta}_1$ zeer waarschijnlijk vertekend is omwille van wegge laten causale variabelen. Bijvoorbeeld: het is zeer waarschijnlijk dat $\log(wage)$ afhangt van het scholingsniveau én dat het scholingsniveau gecorreleerd is met $black$, waardoor $black$ endogeen wordt. Meer specifiek: als zwarte werknemers gemiddeld minder scholing hebben en als scholing een positief effect heeft op het loon, dan is de KK schatter $\hat{\beta}_1$ naar beneden toe vertekend. Kortom, de negatieve schatting $\hat{\beta}_1$ kan te wijten zijn aan wegge laten variabelen en weerspiegelt niet noodzakelijk raciale discriminatie.

4. Het geschat effect van *black* als *exper* = 5 is -0.1305 . Dit betekent dat de voorspelde kans op een hoog loon voor zwarte werknemers met 5 jaar werkervaring 13.05 percentagepunten lager is dan voor niet-zwarte werknemers met 5 jaar werkervaring. De formule is als volgt:

$$\begin{aligned} & \Pr[\widehat{highwage} = 1 | \widehat{black} = 1, exper = 5] - \Pr[\widehat{highwage} = 1 | \widehat{black} = 0, exper = 5] \\ &= \Lambda(\widehat{\beta}_0 + \widehat{\beta}_1 \times 1 + \widehat{\beta}_2 \times 5 + \widehat{\beta}_3 \times 5^2) - \Lambda(\widehat{\beta}_0 + \widehat{\beta}_1 \times 0 + \widehat{\beta}_2 \times 5 + \widehat{\beta}_3 \times 5^2) \\ &= -0.1305. \end{aligned}$$

5.

$$\begin{aligned} \log(\widehat{wage}_{it}) &= \widehat{FE} + \underset{(0.0106)}{0.122} exper_{it} - \underset{(0.000688)}{0.00452} exper_{it}^2 \\ R^2 &= 0.173 \quad SER = 0.352 \end{aligned}$$

6. De dataset is gebalanceerd. Er zijn geen ontbrekende data (voor alle i en t).
7. De regressor $black_{it}$ varieert niet over de tijd: $black_{it} = black_{it'} = black_i$ (voor alle t en t'). Daarom elimineert de within transformatie de regressor $black_{it}$. Het is dus niet mogelijk $black_{it}$ op te nemen in (3). Anders gezegd: het fixed effect α_i absorbeert $black_{it}$.

Vraag 2

1. Ja. Naast *educ* en *exper* zijn er nog tal van variabelen, zoals professionele skills, een goede werkhouding, volharding, etc., die effecten hebben op het loon; al deze variabelen zijn vervat in u in (4). Bovendien zijn al deze variabelen wellicht positief gecorreleerd met *educ*. Daarom is *educ* wellicht gecorreleerd met u en, bijgevolg, is u wellicht endogeen.
2. Een set van instrumenten Z_1, \dots, Z_l moet voldaan aan deze twee voorwaarden:

- (a) de set Z_1, \dots, Z_l is relevant: de $l \times 3$ covariantiematrix

$$\text{cov}(\{Z_1, \dots, Z_l\}, \{educ, exper, exper^2\})$$

heeft rang 3;

- (b) alle Z_1, \dots, Z_l zijn exogeen: $\text{corr}(Z_j, u) = 0$ voor alle j ; dit betekent dat elke Z_j ongecorreleerd moet zijn met de causale determinanten van $\log(wage)$, met uitzondering van $educ$, $exper$ en $exper^2$.

3. *motheduc* en *fatheduc* dienen (samen met *exper*, $exper^2$) te voldoen aan de twee voorwaarden in de vorige vraag.

- (a) We veronderstellen dat, ongeacht het niveau van *exper*, de variabelen *motheduc* en *fatheduc* positief gecorreleerd zijn met *educ*: mensen zijn gemiddeld hoger opgeleid naarmate hun ouders hoger opgeleid zijn. Om dit na te gaan, kunnen we een KK regressie uitvoeren van *educ* op *motheduc*, *fatheduc* en de exogene regressoren in (4):

$$\widehat{educ} = 9.103 + 0.158 \text{motheduc} + 0.190 \text{fatheduc} + 0.0452 \text{exper} - 0.00101 \text{exper}^2.$$

(0.424)
(0.0355)
(0.0324)
(0.0419)
(0.00132)

De coëfficiënten van *motheduc* en *fatheduc* zijn positief (zoals verwacht) en statistisch zeer significant. Er is sterke evidentie dat aan de voorwaarde van relevantie voldaan is.

- (b) De voorwaarde van exogeniteit, met name dat $\text{corr}(\text{motheduc}, u) = 0$ en $\text{corr}(\text{fatheduc}, u) = 0$ kan niet formeel getoetst worden. Er zijn redenen om te twifelen aan die voorwaarde. Het is zeer waarschijnlijk dat de individuele karakteristieken vermeld in Vraag 2.1 (professionele skills, werkhouding, etc.) beïnvloed zijn door *motheduc* en *fatheduc*. Dit veroorzaakt correlatie tussen *motheduc* en *fatheduc* enerzijds en u anderzijds.

4.

$$\widehat{\log(wage)} = 0.0481 + 0.0614 \text{educ} + 0.0442 \text{exper} - 0.000899 \text{exper}^2$$

(0.430)
(0.0333)
(0.0155)
(0.000430)

$R^2 = 0.136 \quad SER = 0.675$

5. $\hat{\beta}_1 = 0.0614$: een stijging van *educ* met 1 jaar doet, gemiddeld genomen, het loon stijgen met $(\hat{\beta}_1 \times 100)\% = 6.14\%$ (onder de voorwaarde van geldige instrumenten).

6. De eerste stap geeft inderdaad sterke indicaties dat aan de relevantievoorwaarde voldaan is: zie deel (a) van het antwoord op vraag 2.3. De geschatte coëfficiënten in de tweede stap zijn identiek aan die in vraag 5; de standaardfouten zijn echter verschillend. De standaardfouten in de tweede stap zijn fout omdat ze gebaseerd zijn op de resttermen uit stap twee:

$$y_i - \hat{x}_i' \hat{\beta}^{2SLS} = y_i - \hat{x}_i' \hat{\beta}^{IV}$$

in plaats van de correcte resttermen

$$\hat{u}_i = y_i - x_i' \hat{\beta}^{IV},$$

die x_i' gebruiken i.p.v. \hat{x}_i' . De standaardfouten in stap twee zijn fout; de standaardfouten in vraag 5 (bekomen met `ivregress 2sls`) zijn juist.