

Econometrie: tussentijdse test 2

Instructies:

- Voor het met werken met Stata mag je gebruik maken van test2.do en dat bestand verder aanvullen.
- Gebruik heteroskedasticiteits-robuste standaardfouten (behalve voor logit/probit).
- Voor hypothesetoetsen, wees zorgvuldig: geef steeds alle stappen aan zoals opgelijst op p. 17 van Lab2.pdf.
- Als je een getal rapporteert, vermeld steeds de eenheid.
- Schrijf je antwoorden neer op het antwoordblad.

Vraag 1

De dataset `wages_panel.dta` bevat gegevens over het uurloon en kenmerken van 545 Amerikaanse werknemers van 1980 tot 1987. Het is een panel dataset met $n = 545$ en $T = 8$. De relevante variabelen zijn:

- **nr**: uniek identificatienummer van de werknemer
- **year**: jaar van de observatie
- **wage**: uurloon in dollar (van de werknemer in dat jaar)
- **exper**: aantal jaren werkervaring (van de werknemer in dat jaar)
- **black**: 0/1 variabele, gelijk aan 1 als de werknemer zwart is

In eerste instantie negeren we de panel structuur van de dataset. Dus we beschouwen elke werknemer in elk jaar als een onafhankelijke waarneming. Schat het model

$$\log(wage_{it}) = \beta_0 + \beta_1 black_{it} + \beta_2 exper_{it} + \beta_3 exper_{it}^2 + u_{it} \quad (1)$$

(d.w.z. zonder fixed effects) met KK. De schattingresultaten hoeven niet gerapporteerd te worden. Hint: gebruik de prefixen `i.` en `c.` voor de regressoren (zie Lab3.pdf, p. 16 en 21, voor voorbeelden).

1. (2 punten) Interpret *SER*. De interpretatie moet concreet zijn: ze moet betrekking hebben op *deze* regressie.
2. (2 punten) Interpreteer het effect van *exper* op het uurloon. Bereken hiervoor (en interpreteer) het marginaal effect van *exper* op het uurloon voor een werner met 5 jaar werkervaring. Hint: gebruik het commando `margins`.
3. (2 punten) Besluit je uit de regressieresultaten dat er loondiscriminatie is tegen zwarte werknemers? Of denk je dat de regressie in (1) ongeschikt is om die vraag te beantwoorden? Motiveer je antwoord.

Maak nu 0/1 variabele *highwage_{it}* aan, met waarde 1 als $\log(wage_{it})$ boven het steekproefgemiddelde is (d.w.z. het steekproefgemiddelde van $\log(wage_{it})$, over alle werknemers en alle jaren), en waarde 0 anders. Schat nu het volgende logit model, terwijl je nog steeds de panel structuur negeert (de schattingresultaten hoeven niet gerapporteerd te worden):

$$\Pr[highwage_{it} = 1 | black_{it}, exper_{it}] = \Lambda(\beta_0 + \beta_1 black_{it} + \beta_2 exper_{it} + \beta_3 exper_{it}^2) \quad (2)$$

Hint: gebruik opnieuw de prefixen `i.` en `c.`

4. (3 punten) Bereken het effect van *black* als *exper* = 5. Gebruik `margins`. Geef een nauwkeurige interpretatie van het getal dat je bekomt. Schrijf ook de formule neer voor dit effect, in termen van model (2).

Hou vanaf hier rekening met de panel structuur van de dataset.

5. (1 punt) Schat het fixed-effects model

$$\log(wage_{it}) = \alpha_i + \beta_1 exper_{it} + \beta_2 exper_{it}^2 + u_{it} \quad (3)$$

en rapporteer de resultaten in standaardvorm.

6. (1 punt) Is de dataset gebalanceerd of ongebalanceerd? Leg kort uit.
7. (2 bonuspunten) Vergelijk (3) met (1): waarom is *black_{it}* niet opgenomen in (3)? Zou het mogelijk zijn *black_{it}* op te nemen in (3)? Waarom (niet)?

Vraag 2

De dataset `wages_iv` bevat cross-sectionele gegevens over 428 Amerikaanse werkende vrouwen in 1975. De relevante variabelen zijn:

- `wage`: uurloon in dollar
- `educ`: aantal jaren scholing (vanaf de leeftijd van zes jaar)
- `exper`: aantal jaren werkervaring
- `motheduc`: aantal jaren scholing van de moeder
- `fatheduc`: aantal jaren schooling van de vader

Beschouw het model

$$\log(\text{wage}_i) = \beta_0 + \beta_1 \text{educ}_i + \beta_2 \text{exper}_i + \beta_3 \text{exper}_i^2 + u_i \quad (4)$$

en veronderstel dat exper en exper^2 exogeen zijn.

1. (2 punten) Zou educ endogeen kunnen zijn in (4)? Waarom (niet)?
2. (2 punten) Wat zijn de formele vereisten voor geldige instrumenten voor (4)? (Kopieer de vereisten *niet* van de cursusslides. Uw antwoord moet concreet toegepast zijn op (4).)
3. (2 punten) Denk je dat motheduc en fatheduc geldige instrumenten kunnen zijn (ter aanvulling van exper en exper^2 , die tegelijk fungeren als regressor en instrument)? Reflecteer apart over de twee voorwaarden voor geldigheid van instrumenten.
4. (1 punt) Schat de IV regressie in (4) met motheduc , fatheduc , exper en exper^2 als instrumenten en rapporteer de resultaten in standaardvorm.
5. (2 punten) Interpreteer de IV schatting van β_1 .
6. (2 bonuspunten) Bereken nu de IV schattingen van β_0 tot β_3 in twee KK stappen:
 - (a) eerste stap: regresseer educ op motheduc , fatheduc , exper en exper^2 met KK en bereken de gefitte waarden

$$\widehat{\text{educ}}_i = \hat{\pi}_0 + \hat{\pi}_1 \text{motheduc}_i + \hat{\pi}_2 \text{fatheduc}_i + \hat{\pi}_3 \text{exper}_i + \hat{\pi}_4 \text{exper}_i^2$$

- (b) tweede stap: regresseer $\log(wage)$ op \widehat{educ} , $exper$, and $exper^2$ met KK.

(Rapporteren van beide stappen hoeft niet.) Bekijk de geschatte regressie in de eerste stap. Zie je indicaties dat er voldaan is aan de relevantievoorwaarde voor de instrumenten (vraag 2.3 hierboven)? Bekijk de regressieresultaten in de tweede stap. Vergelijk de geschatte coëfficiënten en de standaardfouten in stap 2 met de geschatte coëfficiënten en de standaardfouten in vraag 5: wat vind je? Heb je een verklaring waarom de standaardfouten verschillend zijn? Welke zijn juist?