

Uitwerkingen oefentoets Empirische Econometrie

Vraag 1

- a. Een voordeel van de gepoolde regressie ten opzichte van afzonderlijke cross-sectie regressies is dat je meer waarnemingen hebt, dus meer informatie en dus nauwkeuriger schattingen. Ander voordeel is dat de onderzoeksvraag met name een tijdsdimensie heeft. Hoeveel meer gaan landen met elkaar handelen als gevolg van de invoering van een gezamenlijke munteenheid? Een mogelijk nadeel is dat de gepoolde regressie veronderstelt dat regressiecoëfficiënten constant blijven over de tijd. Als zich echter structurele breuken voordoen, dan bevat het gepoolde model een specificatiefout.
- b. Fixed effects impliceert dat we α_{ij} beschouwen als onbekende regressiecoëfficiënten bij dummyvariabelen, die we toevoegen aan de regressie. Random effects impliceert dat we α_{ij} als onderdeel van de storingsterm beschouwen. Grote verschil is dus dat bij fixed effects α_{ij} niet stochastisch is, terwijl bij random effects α_{ij} een kansvariabele is.
- c. Een voordeel van fixed effects is dat we minder statistische aannames nodig hebben om consistente schatters te krijgen. Bij random effects moeten we bijvoorbeeld aannemen dat regressoren strict exogeen zijn ten opzichte van het random effect.
- d. Een voordeel van random effects is dat veel minder onbekende regressiecoëfficiënten geschat hoeven te worden ten opzichte van fixed effects. Dat leidt tot een schatter met kleinere standaardfouten.
- e. Glick & Rose (2002) gebruiken een fixed effects specificatie om dezelfde reden als in onderdeel c. Bovendien hebben ze zoveel gegevens dat ook fixed effects schatters nauwkeurig genoeg zijn.

Vraag 2

- a. Scholing is een endogene variabele in de loonvergelijking. Verschillende oorzaken kunnen hieraan ten grondslag liggen, maar Angrist & Krueger (1991) wijzen met name op het gevaar van weggelaten relevante factoren in de loonvergelijking. Wanneer een relevante verklarende variabele ten onrechte wordt weggelaten en bovendien gecorreleerd is met de regressoren dan ontstaat er endogeniteit d.w.z. afhankelijkheid tussen regressoren en storingsterm. In het geval van de loonvergelijking is talent een belangrijke weggelaten variabele. Talent is bovendien hoogstwaarschijnlijk gecorreleerd met scholing.
- b. We schatten

$$x_i = \pi_1 + \pi_2 z_i + v_i,$$

met kleinste kwadraten. De residuele kwadratensom is gelijk aan:

$$\sum_i \hat{v}_i^2 = (x_i - \hat{\pi}_1 - \hat{\pi}_2 z_i)^2.$$

De partiele afgeleiden naar $\hat{\pi}_1$ en $\hat{\pi}_2$ zijn:

$$\frac{\partial \sum \hat{v}_i^2}{\partial \hat{\pi}_1} = -2 \sum_i (x_i - \hat{\pi}_1 - \hat{\pi}_2 z_i)$$

$$\frac{\partial \sum \hat{v}_i^2}{\partial \hat{\pi}_2} = -2 \sum_i z_i (x_i - \hat{\pi}_1 - \hat{\pi}_2 z_i)$$

Beide afgeleiden gelijk aan nul stellen geeft de normaalvergelijkingen:

$$\sum_i (x_i - \hat{\pi}_1 - \hat{\pi}_2 z_i) = 0$$

$$\sum_i z_i (x_i - \hat{\pi}_1 - \hat{\pi}_2 z_i) = 0$$

c. De normaalvergelijkingen kunnen geschreven worden als:

$$\sum x_i = \hat{\pi}_1 \sum 1 + \hat{\pi}_2 \sum z_i,$$

$$\sum z_i x_i = \hat{\pi}_1 \sum z_i + \hat{\pi}_2 \sum z_i^2.$$

Omdat z een 0-1 variabele is geldt $\sum z_i = \sum z_i^2$. Wanneer we het verschil van de normaalvergelijkingen nemen krijgen we:

$$\sum (1 - z_i) x_i = \hat{\pi}_1 \sum (1 - z_i) + 0 \text{ en dus } \hat{\pi}_1 = \frac{\sum (1 - z_i) x_i}{\sum (1 - z_i)}.$$

Tevens leiden we af uit de tweede normaalvergelijking dat:

$$\hat{\pi}_1 + \hat{\pi}_2 = \frac{\sum z_i x_i}{\sum z_i} \text{ en dus } \hat{\pi}_2 = \frac{\sum z_i x_i}{\sum z_i} - \frac{\sum (1 - z_i) x_i}{\sum (1 - z_i)}.$$

Verdelen we volgens de dummy variabele z de gegevens in twee groepen (1 en 2) dan zijn $\frac{\sum z_i x_i}{\sum z_i} = \bar{x}_2$ en $\frac{\sum (1 - z_i) x_i}{\sum (1 - z_i)} = \bar{x}_1$ de gemiddelden van die twee groepen.

Dus we kunnen schrijven $\hat{\pi}_2 = \bar{x}_2 - \bar{x}_1$ oftewel de schatter is simpelweg het verschil in gemiddelden.

d. We kunnen schrijven:

$$\hat{\beta} = \frac{\sum_i (z_i - \bar{z})(y_i - \bar{y})}{\sum_i (z_i - \bar{z})(x_i - \bar{x})} = \frac{\sum_i (z_i - \bar{z})^2 \sum_i (z_i - \bar{z})(y_i - \bar{y})}{\sum_i (z_i - \bar{z})^2 \sum_i (z_i - \bar{z})(x_i - \bar{x})}.$$

dus de IV schatter van β is gelijk is aan de hellingscoefficient van de regressie van y op een constante en z gedeeld door de hellingscoefficient van de regressie van x op een constante en z . Gegeven het resultaat in onderdeel c. kunnen we de IV schatter ook schrijven als

$$\hat{\beta} = \frac{\bar{y}_2 - \bar{y}_1}{\bar{x}_2 - \bar{x}_1}.$$

In woorden is de Wald schatter dus het gemiddelde verschil tussen de 2 groepen in de uitkomstvariabele gedeeld door gemiddeld verschil in de endogene regressor.

Vraag 3

Ten eerste is volgens de modelspecificatie de dividend-prijs ratio niet een exogene regressor. Hoewel x_t ongecorrleerd is met de huidige storingsterm u_t (en ook toekomstige storingstermen) is de regressor wel gecorreleerd met u_{t-1} (via v_{t-1}). Dus de kleinste kwadratenschatter van β zal onzuiver zijn en de t-toetsgrootheid zal niet langer t-verdeeld zijn onder de nulhypothese dat $\beta=0$.

Ten tweede heeft de dividend-prijs ratio waarschijnlijk een stochastische trend d.w.z. $\phi=1$, terwijl aandelenrendement stationair is. Wanneer dit het geval is zal de t-toets ook niet langer t-verdeeld zijn en is er sprake van een schijnregressie.