

Oppgave 1.

a) Flere løsninger er mulig, men effektivt å ta utgangspunkt i en ligning ala (3) i Bratsberg og Raaum (2012) og diskutere identifikasjonsutfordringer når ulike typer data er tilgjengelig. Sentralt å få fram utfordringen med mulig endogen seleksjon av norske ansatte når innvandrersandelen endres, se diskusjon i Bratsberg og Raaum s.1182. Kandidatene bør få fram problemene med identifikasjon dersom bare tverrsnittsdata eller pooled cross-section-time series data er tilgjengelig og hvordan problemet reduseres ved ekte paneldata (arbeidere følges over tid) ved bruk av faste individeffekter. Effektive kandidater bruker diskusjonen på denne deloppgaven også i besvarelsen på deloppgave b).

b) Bør kommentere at kolonne (1) uten faste individeffekter lider under problemet med endogen seleksjon av norske arbeidere (trekker seg ut av arbeidsmarkedet eller skifter sektor) ved innvandring av arbeidskraft, se også Bratsberg og Raaum s. 1190. Hvis norske arbeidere med lav lønnsnivå systematisk trekker seg ut av arbeidsmarkedet når innvandrerne strømmer inn, vil en relasjon som estimert i kolonne (1) undervurdere lønnsnedgangen for norske arbeidere som følge av innvandringen. Effektive kandidater vil benytte svaret på a) i denne diskusjonen. Få fram at problemet reduseres ved individ-fixed effects-spesifikasjonen og spesifikasjonene (3) og (4) der bare arbeidere som var i hver av undersektorene hele perioden inngår.

Konfidensintervallet rundt koeffisienten i kol (2) er (-1.128, -0.32). Estimerte koeffisientene i kol (3) og (4) ligger begge klart innenfor dette intervallet.

c) Rett fram regning gir at effekten av et prosentpoengs økning i M/N på prosentvis lønnsvekst for norske arbeidere er gitt ved

$$d \ln W / d(M/N) = \theta \cdot 1 / (1 + M/N) = \theta \cdot (N / (N + M)) = \theta \cdot (1 - M / (N + M))$$

der θ er koeffisienten foran $\ln(1 + M/N)$ i modellen.

Estimatet -0.570 og evaluert i gjennomsnittsverdi på $(M/N + M) = 0.08$ innebærer dermed at 1%-poeng økning i M/N gir 0.52 % lønnsreduksjon.

Rett fram regning (differensiering) gir at elastisiteten av lønna mhp M er

$$(dW/dM)(M/W) = \theta \cdot 1 / (1 + M/N) \cdot M/N = \theta \cdot M / (M + N)$$

Estimatet -0.57 og evaluert i gjennomsnittsverdi av $M/(M+N)=0.08$ gir elasticitet på -0.046

Innebærer at 5% økning i innvandrerdelen M reduserer lønnsveksten med 0.23%.

d) Norske og utenlandske arbeidere perfekte substitutter innebærer at $\sigma_M \rightarrow \infty$ og koeffisienten foran $\ln(1+M/N)$ blir dermed $\theta = -1/\sigma_j$ og at det gir estimat på substitusjonselasticiteten mellom aktivitetene på $\widehat{\sigma}_j = -\frac{1}{\hat{\theta}}$. Estimert koeffisient på -0.724

innebærer under denne forutsetningen en estimert substitusjonselasticitet $\widehat{\sigma}_j = 1.4$.

e) Inkludering av aktivitetsspesifikk dummy interagert med år vil være perfekt korrelert med $\ln(1+M/M+N)$ -variabelen som bare varierer over aktivitet og år, og i en sånn spesifikasjon vil ikke effekten av interessevariabelen la seg identifisere.

f) test om $\theta = -1$ med vanlig t-test Testobservatoren blir rett fram $(-0.724 - (-1))/0.202 = 1.37$. Kan ikke forkaste nullhypotesen om $\theta = -1$. Testobservatoren for estimatet fra kolonne (4) blir 2.33. Forkaster nullhypotesen der.

Oppgave 2.

a) Her er det rett fram sharp RDD (Finland) og fuzzy RDD (Sverige)

Bør formulere eksplisitt modellene som svarer til hver av tilfellene slik det er gjort i Petterson-Lidbom sin artikkel. Få fram forskjellen mellom Sharp og Fuzzy RDD. Få fram viktigheten av å inkludere effekt av og spesifisere polynomfunksjon i «forcing»-variabelen.

I undersøkelse av validiteten av strategiene bør inngå diskusjon av:

1) Balanseringstester og relasjon med og uten kontroller. Kandidatene bør få fram at troverdig identifikasjon ved RDD innebærer at inkludering av kontroller bare påvirker skarphet av estimater (estimerte standardavvik for interessekoeffisienten) mens estimatene er upåvirka.

2) global vs lokal identifikasjon (observasjoner i omegn av cutoffene)

3) Standard test av 1.steget i fuzzy RDD

b) Kandidatene bør innse at i dette tilfellet er forskeren ikke i stand til å separere effekten av kommunestyrestørrelse og effekten av statstilskudd på kommunale utgifter og et RDD-estimat vil representere den samlede effekten av de to variablene på kommunale utgifter.

Oppgave 3.

Bør få fram at Kontrollert eksperiment i utgangspunktet er en ideell metode for å estimere kausal effekt av klassestørrelse og som under ideelle betingelser kan gi konsistent anslag på effekten ved å beregne differansen i gjennomsnittlige elevprestasjoner mellom elever tilfeldig plassert i små (treatment) versus store klasser. Relevant å kommentere problemer som kan oppstå i praksis i kontrollerte eksperimenter av denne type som bl.a avvik fra tilfeldig fordeling av elever i utgangspunktet. Kan dessuten tenkes at noen elever bytter gjennom skoleåret fra store til små klasser pga atferdsproblemer etc. Bør også diskutere problem knytta til at selve eksperimentsituasjonen kan påvirke læreratferden-Hawthorneffekten og «John Henry»-effekten som er behandlet i artikkelen til Krueger section E s. 526-27 og dermed komplisere tolkningen av en eventuell estimert klassestørrelseeffekt basert på RCE-oppsettet.

Relevant å beskrive hvordan man kan undersøke om initialplasseringen av elevene i små vs. store klasser faktisk var tilfeldig. Sammenligning av gjennomsnittsverdier på elevkarakteristika mellom små og store klasser, evt teste ved enkel regresjon mellom dummy for små vs stor klasse mot elevkarakteristika hver for seg og teste ved t-test om f.eks alder er forskjellig mellom små og store klasser. Mer generell test oppnås ved å kjøre en regresjon mellom dummy for små vs. stor klasse mot alle elevkjennetegnene (alder, indikator for kjønn, indikator for høyere utdanning, indikator for innvandrere). Faktisk randomisering svarer til nullhypotese om at alle koeffisientene foran elevkjennetegnene er lik null i en slik regresjon. Kan testes ved standard F-test.

Alternativt kan det estimeres en lineær regresjonsmodell med resultat på lestesten som avhengig variabel med dummy for liten klasse (treatmentdummy) som forklaringsvariabel med og uten kontroll for elevkarakteristika med OLS. Faktisk tilfeldig fordeling av elever tilsier at estimert koeffisient foran treatmentdummien bør være like i modell med og uten kontrollvariable (men inkludering av elevkjennetegn kan redusere estimert standardavvik til koeffisienten foran treatmentdummien pga redusert restleddsvarians)

b) Kandidatene bør på grunnlag av informasjonen kunne formulere en bakenforliggende modell av samme type som modellen s. 517 i Krueger (1999):

$$(3) \quad CS_{jos} = \pi_0 + \pi_1 S_{jos} + \pi_2 R_{jos} + \pi_3 X_{jos} + \delta_s + \tau_{jos}$$

$$(4) \quad Y_{jos} = \beta_0 + \beta_1 CS_{jos} + \beta_2 X_{jos} + \alpha_s + \epsilon_{jos}$$

Her er CS faktisk klassestørrelse, S er dummy for liten klasse, R er dummy for regulær klasse, X er vektor av elevkarakteristika, Y er prestasjoner. (3) er førstestegsrelasjon, (4) er strukturligning. Begrunnelse: Elevmobilitet mellom skoler og klasser, samt at klassestørrelsen varierte innen S og R-klasser. Gjerne henvis til diskusjonen i spm a) her. Standard forklaring av IV-2SLS-estimering og at instrumentrelevans kan undersøkes ved F-test på $H_0: \pi_1 = \pi_2 = 0$ i (3). Her er også relevant å foreslå å estimere modell med og uten elevkjenntegn. Hvis troverdig identifikasjon så bør estimat på β_1 ikke være systematisk forskjellig i de to spesifikasjonene.

10 elever ekstra i klassen reduserer scoren med knapt 9 prosentiler i 1.trinn ifølge 2SLS-estimatene.