

## Sensorveiledning SØK3514 Vår 2020

Dette er en relativt åpen oppgave der altså alle hjelpemidler vil være tilgjengelig ved hjemmeeksamen. Det bør derfor legges vekt på god diskusjon av relevante momenter. Under spørsmål a) forventes at kandidatene utnytter stoff fra pensumartiklene av Angrist and Lavy og Krueger mens stoff fra pensumartikkelen av Becker, Hornung og Woessman er sentralt ved svar på spørsmål b).

Økonometriske utfordringer under spørsmål a) gjelder generelt endogen klassestørrelse og endogen elevsammensetning. Diskusjon av utelatte variable og særlig problemer knyttet til sortering vil være relevant. Konkretisering av sorteringsmekanismer (plassering av svake elever i små klasser, flytting mellom skoler, lærerstorting mellom klasser og skoler) honoreres.

Når det gjelder empirisk strategi bør det gjøres presist rede for hvordan Angrist og Lavy utnytter klassesdelingsregelen i et «fuzzy discontinuity design». Hovedpoenget er altså at klassesdelingsregelen gir en kilde til eksogen variasjon og at predikert klassestørrelse benyttes som instrument for faktisk klassestørrelse. I forbindelse med denne hovedstrategien er det viktig å diskutere hvorfor det bør kontrolleres for kullstørrelse samt fordeler og ulemper ved å bruke hele utvalget eller kun observasjoner nær diskontinuitetene. Det bør også gjøres rede for hvilke andre variable det kontrolleres for og hvilke sensitivitetsanalyser som gjennomføres, bl.a. analyser av potensiell heterogenitet.

Videre bør det gjøres rede for hvordan Krueger utnytter eksperimentdata. Et naturlig utgangspunkt vil her være at dersom vi har et perfekt randomisert eksperiment der elever og lærere plasseres i ulike klassetyper, kunne en rett fram teste om gjennomsnittlige elevprestasjoner er ulike mellom ulike klassetyper. Det bør her gjøres rede for hvordan eksperimentet er designet. Videre bør ulike forhold som gir avvik fra den ideelle eksperimentsituasjonen diskuteres og hvordan forfatteren forsøker å håndtere potensielle problemer med dette skaper. Denne diskusjonen bør konkret knyttes til ulike sensitivitetsanalyser som gjennomføres i artikkelen: i) inkludering av flere kontrollvariable (ved perfekt randomisering skal ikke dette ha betydning), ii) bruk av dummyvariable for initial plassering for å ta hensyn til flytting av elever mellom klassetyper, iii) bruk av «imputed test score» for å ta hensyn til ikke-tilfeldig frafall.

I artikkelen gjennomføres også analyser der faktisk klassestørrelse instrumenteres med klassetypedummyer. Videre gjennomføres analyser for å avdekke kumulative effekter og heterogenitet. Videre diskuterer forfatteren potensielle problemer knyttet til Hawthorne og John Henry effekten. Gode besvarelser bør i hvert fall ha med noen av disse sensitivitetsanalysene.

Den viktigste utfordringen under spørsmål b) gjelder simultanitet mellom utdanningsnivå og mål på industrialisering (kan her argumentere for at simultanitetsskjevheter i OLS-estimatene kan gå begge veier). I tillegg kan en argumentere for at svært mange faktorer er av betydning for industrialisering slik at utelatte variable er et potensielt alvorlig problem.

I forbindelse med diskusjon av den empiriske strategien i artikkelen til Becker, Hornung og Woessmann bør det gjøres rede for hva slags type data som utnyttes: Regionale data innenfor Preussen for utdanningsnivå, mål på industrialisering og kontrollvariable på ulike tidspunkt / perioder. Det kan da argumenteres for at de ulike regionene ikke er så forskjellige i andre

henseender hvilket reduserer potensielle problemer med heterogenitet / utelatte variable. Videre bør det gjøres rede for hovedstrategien i artikkelen som er å utnytte variasjoner i initialt utdanningsnivå som instrument for senere utdanningsnivå som tenkes å påvirke industrialiseringsprosessen. Diskuter her styrken på instrumentet med henvisning til analysen i artikkelen og sammenlign IV- og OLS-estimatene.

Videre bør det gjøres rede for en del robusthetssjekker som gjennomføres i artikkelen. Dette gjelder inkludering av ytterligere kontrollvariable (1816) som kan tenkes å være korrelert med instrumentet - tidlig utdanningsnivå (1816). Videre gjennomføres analyser ved bruk av data for ulike faser i industrialiseringsprosessen det det bl.a. betinges med hensyn på tidlig industrialisering og utnyttes paneldata.

Generelt: Siden pensumartiklene er tilgjengelige under eksamen vil det være naturlig at sentrale resultater diskuteres og særlig om endret spesifisering og/eller utvalg påvirker resultatene i særlig grad.