

Sensorveiledning SØK3514 / SØK8614, Eksamen H 2019

Oppgave 1

a) Forfatterne ønsker å teste om utdanningsnivået har betydning for industrialisering – og da i første rekke ved at teknologioverføring fra England er lettere å få til i områder med bedre utdannet befolkning. Økonometriske utfordringer i den forbindelse er i første rekke simultanitet og potensielt utelatte variable. Når det gjelder simultanitet kan det argumenteres for at industrialisering i første fase øker etterspørselen etter ufaglært arbeidskraft og dermed har en negativ effekt på utdanningsnivået. På den annen side vil industrialisering via økt inntekt kunne øke etterspørselen etter utdanning. Kandidatene kan her formulere en simultan modell for industrialisering og utdanning og diskutere formelt hvordan omvendt kausalitet vil påvirke skjevhet i OLS-estimatorene avhengig av fortegn på feedback. God og presis verbal diskusjon er imidlertid en likeverdig strategi. Når det gjelder potensielle problem med utelatte variable kan en argumentere for at det er mange forhold som påvirker utdanningsnivået og at slike forhold også kan tenkes å påvirke industrialiseringen.

Forfatterne bruker data for ulike regioner innenfor Preussen. Siden en kan argumentere for at regionene innenfor Preussen er relativt like i andre henseender (sammenlignet med ulike land) reduseres problemet med utelatte variable. Hovedstrategien er imidlertid å benytte instrumentvariabelmetoden der utdanningsnivået i 1849 (og senere 1871) instrumenteres ved bruk av mål på utdanningsnivået i 1816. I den forbindelse bør kandidatene kort forklare IV-metoden og hvilke forutsetninger den bygger på. I artikkelen vises at utdanningsnivået i 1816 varierer veldig mye mellom regioner, og hvis disse forskjellene er persistente har vi et relevant instrument (hvilket klart framkommer i de neste delspørsmålene).

b) Table 1 viser estimert effekt av utdanningsnivået 1849 på målet på industrialisering i 1849. Kolonne (1) rapporterer OLS-estimatet som er positivt og statistisk signifikant. Her kan det argumenteres for skjevhet pga simultanitet og potensielt utelatte variable. I kolonne 3 rapporteres IV-estimatet der utdanningsnivået betraktes som endogent og instrumenteres. Selv om estimert effekt reduseres noe er ikke forskjellen fra OLS-estimatet veldig stor. Kolonne 2 rapporterer den viktigste informasjonen angående den tilhørende førstestegsregresjonen. Det bør påpekes at utdanningsnivået i 1816 inngår svært signifikant i førstestegsregresjonen og at F-verdien på testen av styrke / relevans er over 6000 så instrumentet kan karakteriseres som supersterkt

Siden forfatterne kun bruker en instrumentvariabel kan en ikke teste validiteten av instrumentet. Et potensielt problem er at forhold som påvirker utdanningsnivået i 1816 også har betydning for industrialiseringen. I så fall har vi utelatte variable hvis effekt inngår i restleddet og som er korrelert med instrumentet. Dette er bakgrunnen for sensitivitetssjekken i kolonne 4 og 5 der forfatterne inkluderer et sett av kontrollvariable datert 1816. Disse inngår selvsagt i førstestegsregresjonen 4 der vi fortsatt merker oss svært høy styrke på instrumentet. Inkludering av kontrollvariable datert 1816 øker estimert effekt av utdanning i IV-regresjonen i kolonne 5 og kan sies å styrke hypotesen om utdanningsnivåets betydning for industrialisering.

c) Her bør det påpekes at forfatterne inkluderer nivå på industrialisering 1849 som eksogen variabel i modellen for industrialisering 1882 og at denne har stor betydning. Videre ser vi at utdanningsnivået har separat positiv og signifikant effekt. En kan derfor si at

utdanningsnivået både har en indirekte effekt via industrialisering i fase 1 og en direkte effekt på industrialiseringen i fase 2. Fra førstestegsregresjonen i (6) ser vi videre at instrumentet fortsatt har høy styrke (F-verdi lik 65).

Oppgave 2

a) Hovedsak her: Interessevariabelen påvirkes diskontinuerlig av «assignment» / «forcing» variabel. Diskontinuerlige sprang i interessevariabelen gir eksogen variasjon rundt «cut-off» / terskelverdier som utnyttes når en ønsker å estimere kausaleffekten av denne på outcome. Poeng å undersøke særskilt om outcome påvirkes nær cutoff ved bruk av ulike vindusbredder. Argumenter for at «treatment» og «kontroll» sannsynligvis er mer like i andre henseender like over og like under cutoff. Videre poeng er at assignmentvariabelen i de fleste tilfeller må antas å påvirke outcome. Derfor viktig å kontrollere for denne, og å ta hensyn til potensielle ikke-lineariteter (ofte ved polynom) i kontrollfunksjonen. Generelt argument: Interessevariabelen er diskontinuerlig mens andre variable som oftest er kontinuerlige. Kan nevnes at en bør sjekke dette: Er det andre variable som også viser diskontinuitet i omegnen av terskelverdiene for interessevariabelen? I så fall alvorlig utfordring for RDD. Kan også ta opp potensielle problem dersom cut-off kan manipuleres.

«Sharp design»: Interessevariabelen er en deterministisk funksjon (og bestemmes entydig) av assignmentvariabelen. «Fuzzy design»: Interessevariabelen betraktes som stokastisk, men ssh for diskontinuitet påvirkes av assignmentvariabelen nær potensielle cut-offs. Bør nevnes at ved fuzzy benyttes en 2stegs metode der interessevariabelen instrumenteres.

b) For Finland er antall representanter eksakt gitt ved lov og øker sprangvis med folketall. Vi har da en deterministisk funksjon av innbyggertall så her har vi sharp design. For Sverige gjelder minimumskrav for antall representanter innenfor oppgitte intervall. Her har vi derfor fuzzy design.

Grunnmodellen som estimeres – både for Finland og Sverige er gitt ved:

$$(1) Y_{it} = \alpha_i + \lambda_t + \beta Csize_{it} + f(x_{it}) + u_{it}$$

der y er log til per capita kommunale utgifter og $Csize$ log til antall representanter.

Kommenter at det inkluderes kommunespesifikke faste effekter, faste tidseffekter samt ulike polynom på kontrollfunksjonen, det siste dels for robusthetssjekker.

For Finland estimeres (1) direkte ved bruk av OLS, dels basert på alle observasjoner, dels ved bruk av observasjoner innenfor ulike vindusbredder nær terskelverdiene.

For Sverige instrumenteres $Csize$ ved å inkludere et sett av dummyvariable for intervallene i førstestegsregresjonen. Andresteget estimerer (1) med «predikert $Csize$ i stedet for faktisk $Csize$). Første steget inkluderer også kommunespesifikke faste effekter, tidseffekter og ulike polynom på kontrollfunksjonen.

Kan nevnes at en fordel med å studere kommuner innen et land er at samme institusjonelle forhold gjelder, relativt like kulturelle forhold mellom kommuner etc. Dette i motsetning til tverrsnittstudier basert på ulike land.

For Finland finner forfatteren signifikant negativ effekt av $Csize$ og resultatene er relativt robuste. For Sverige er estimert effekt også negativ, men varierer mye med spesifisering og

er heller ikke signifikante. Det er her relevant å nevne at resultatene fra førstestegsregresjonen tilsier at instrumentene ikke er veldig sterke.

c) Artikler på pensum som benytter RDD er Petterson-Lidbom (2008) som studerer effekt på ulike kommunale utfallsvariable av hvilken politisk blokk som har flertall i Svenske kommuner. Assignmentvariabelen er stemmeandel til venstresiden og interessevariabelen er dummy for sosialistisk flertall (som altså gjør en sprang fra 0 til 1 ved stemmeandel på 50 %). Her har vi et eksempel på sharp design. Videre benytter Angrist og Lavy regler for maksimal klassestørrelse i en studie av sammenhengen mellom klassestørrelse og elevprestasjoner. Klasedelingsregelen benyttes for å konstruere et instrument der faktisk klassestørrelse instrumenteres i strukturrelasjonen. Også andre eksempler kan være relevante. Det viktigste er å få fram hva som er outcomevariabel, forklaringsvariabel av interesse, hva som gir diskontinuitet i denne og om vi har sharp eller fuzzy design.

Oppgave 3

a) I mange studier av effekt av økt arbeidsinnvandring på lønn tar en gjerne utgangspunkt i at eksogene hendelser utenfor landet gir opphav til økt innstrømming av arbeidskraft. Dette gjelder også artikkelen til Bratsberg og Raaum der utvidelse av EU samt lavkonjunktur i Svensk anleggssektor ga økt innstrømming av arbeidskraft til Norge. Men selv om mye av innstrømmingen skyldes forhold utenfor Norge er det likevel flere økonometriske problemer:

For det første at utenlandsk arbeidskraft naturlig nok strømmer inn i gode tider og sorterer seg til regioner, bransjer og yrker hvor det er mangel på arbeidskraft. Da er det jo i utgangspunktet høy lønnsvekst i disse segmentene. Det forfatterne gjør for å møte disse utfordringene er å bruke data for ulike aktiviteter innen bygg og anlegg. En kan argumentere for at konjunktursituasjonen er ganske lik mellom aktiviteter innenfor B&A. Videre utnytter de strenge krav til sertifisering for enkelte aktiviteter som rørlegger, elektriker, brobygging etc. Her kan en anta at innstrømmingen av arbeidskraft fra Polen og Baltikum har svært liten betydning. Disse aktivitetene utgjør da en kontrollgruppe i studien. På den annen side er det ingen tilsvarende sertifiseringsregler i mange andre aktiviteter (malere, snekkere etc) som da utgjør «behandlingsgruppen». En annen utfordring er at innfødte trekker seg ut fra aktiviteter / områder med høy arbeidsinnvandring, og de som gjør det er neppe et tilfeldig utvalg. Hvordan forfatterne håndterer dette kan med fordel diskuteres under b)

b) Tabellen viser at estimert effekt av økt innvandrerandel varierer mye med spesifisering (uten, med individspesifikke effekter), og mellom ulike utvalg. Når de bruker «balansert panel» droppes observasjoner for norske arbeidere som starter i løpet av perioden, og norske arbeidere som slutter / trekker seg ut. Og når de dropper exit er det liten forskjell i estimert effekt fra balansert panel og også når de inkluderer individspesifikke faste effekter. Det følger da at disse estimatene er mest troverdige, men hvilke av disse som bør foretrekkes er imidlertid ikke helt åpenbart. Hovedsaken er at inkludering av individspesifikke effekter i (2) kontrollerer for uobserverbar heterogenitet som tar hensyn til endogen sortering av innenlandske arbeidere. I (3), (4) og (6) inkluderes ikke personer som endogent trekker seg ut av vedkommende aktivitet (pga økt innvandrerandel) slik at en derfor synes å unngå problemet med endogen sortering.

c) Her er det lurt å tenke på hva som blir basiskategori, nemlig norske arbeidere som er sysselsatt i samme aktivitet / delsektor hele perioden. En kan da argumentere for at estimert effekt for basis vil bli lik de forfatterne rapporterer i kolonne 3 og 4 (hvis individspesifikke effekter inkluderes). Ellers er dette et spørsmål der gode resonnerer honorerer.