

Instrumentvariabler – en introduksjon for samfunnsforskere

Nicolai T. Borgen

Institutt for sosiologi og samfunnsgeografi, Universitet i Oslo
nicolai.borgen@sosgeo.uio.no

En vanlig tilnærming for å identifisere kausale effekter i samfunnsvitenskapelig forskning er kontrollvariabelmetoden. Ofte er imidlertid én eller flere bakenforliggende variabler uobserverte, det er målefeil i variablene og/eller det er resiprok kausalitet. I disse tilfellene vil kontrollvariabelmetoden gi skjeve estimater. I denne artikkelen presenterer og diskuterer jeg en metode som kan benyttes for å unngå skjevhet som skyldes uobserverte variabler, målefeil og/eller resiprok kausalitet: instrumentvariabelmetoden. Målet med artikkelen er å gjøre leseren i stand til å vurdere validiteten av resultater i studier som bruker instrumentvariabler og identifisere variabler som kan benyttes som instrumenter i egen forskning. Diskusjonen vil vise at instrumentvariabelmetoden i teorien kan benyttes for å unngå skjevhet som skyldes korrelasjon mellom restleddet og forklaringsvariabelen, og derfor er et nyttig verktøy for å identifisere kausaleffekter. Det er imidlertid vanskelig å finne instrumenter som oppfyller alle forutsetningene i instrumentvariabelmetoden, og konsekvensene av brudd på forutsetningene er potensielt store. Et viktig poeng i artikkelen er derfor at vellykket bruk av instrumentvariabler krever en kreativ forsker som har en god forståelse av forskningsområdet.

Nøkkelord: Instrumentvariabelmetoden, instrumentvaliditet, instrumentrelevans, svake instrumenter, LATE, monotonitet.

Innledning¹

En stor del av den empiriske samfunnsforskningen handler om årsaks-sammenhenger, men evnen til å trekke slutninger om disse sammenhengene

begrenses av tilgangen til gode data (Morgan og Winship 2007: 1). Dette skyldes at vi i all hovedsak må benytte observasjonsdata i samfunnsforskning. I randomiserte eksperimenter, som av mange fremheves som gullstandarden for kausal identifisering, vil forskeren ha full kontroll over hvem som skal motta en påvirkning og når denne påvirkningen mottas. Ved bruk av observasjonsdata mister forskeren denne kontrollen, og det blir derfor vanskeligere å fastslå både hva styrken på sammenhengene er, og i hvilken retning årsakssammenhengene går. I tillegg er det en stor utfordring ved observasjonsdata at variablene ikke er målt perfekt. Noen ganger skyldes målefeil feilrapportering og feilregistrering, mens andre ganger oppstår målefeil fordi vi ønsker å måle abstrakte begreper, som holdninger eller kognitive ferdigheter.

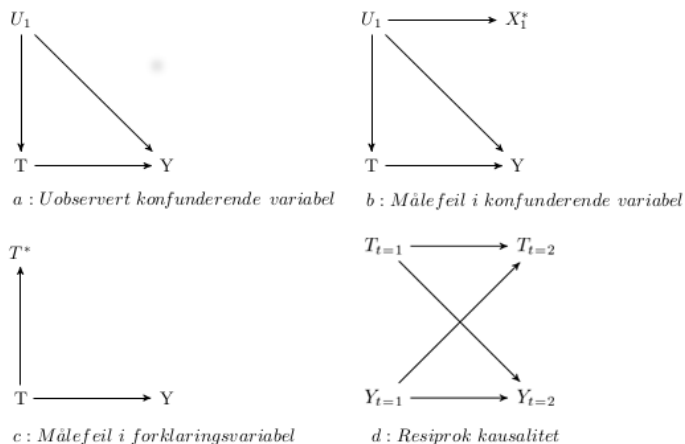
Den vanligste strategien for å identifisere kausale effekter med observasjonsdata er kontrollvariabelmetoden. Denne metoden går primært ut på å kontrollere for observerte konfunderende variabler. Instrumentvariabelmetoden representerer en helt annen måte å løse problemene med kausal identifisering på, og utgjør et alternativ til kontrollvariabelmetoden når de konfunderende variablene er uobserverte, det er målefeil i de konfunderende variablene, det er målefeil i forklaringsvariabelen og/eller når det er resiprok kausalitet. Instrumentvariabelmetoden er derfor et nyttig verktøy for å identifisere kausale effekter med observasjonsdata. Målet med artikkelen er å gjøre leseren i stand til å vurdere validiteten av resultater i studier som bruker instrumentvariabler og identifisere variabler som kan benyttes som instrumenter i egen forskning. For lesere som ønsker selv å benytte metoden, bør artikkelen suppleres med flere bidrag. Morgan og Winship (2007) gir en intuitiv presentasjon, og inkluderer flere eksempler enn denne artikkelen. Se også Bollen (2012), Angrist et al. (1996), Angrist og Pischke (2009), Wooldridge (2006) og Wooldridge (2010).

Jeg legger til grunn at leseren er kjent med kontrollvariabelmetoden (Skog 2004) og kausale grafer (Pearl 2009). For en intuitiv presentasjon av kausale grafer, se for eksempel Mohn og Hermansen (2013) i dette nummeret av Sosiologi i dag. Jeg starter artikkelen med å beskrive noen situasjoner hvor vi bør benytte instrumentvariabler, før jeg introduserer instrumentvariabelmetoden. Deretter vil jeg diskutere noen fallgruver når vi benytter instrumentvariabelmetoden. Jeg diskuterer problemer knyttet til invalide instrumenter, irrelevante instrumenter, svake instrumenter, effektheterogenitet, monotonitet, antall instrumenter og bruk av kontrollvariabler. Et viktig poeng som tydeliggjøres gjennom denne diskusjonen er at den største utfordringen når vi benytter

instrumentvariabler ikke er statistisk eller teknisk, men teoretisk. Vellykket bruk av instrumentvariabler avhenger av at forskeren er kreativ og har en god forståelse av forskningsområdet (Angrist og Pischke 2009: 117). Jeg kommer derfor i liten grad inn på den praktiske gjennomføringen av instrumentvariabelmetoden. For brukere av Stata gir Cameron og Trivedi (2010) en god innføring i det praktiske knyttet til estimering og testing. Videre vil jeg begrense meg til å gjennomgå instrumentvariabelmetoden i lineær regresjon, ettersom instrumentvariabelmetoden er mer komplisert i ikke-lineær regresjon. Se Bollen (2012: 47–50) for en oversikt over mulige tilnærminger når den avhengige variabelen er kategorisk.² For en beskrivelse av instrumentvariabler i kvantilregresjon, se Angrist og Pischke (2009: 283–291).

Hvorfor benytte instrumentvariabler?

Når vi benytter observasjonsdata vil kontrollvariabelmetoden ofte gi et feilaktig bilde av sammenhengen mellom variabler. Figur 1 gir eksempler på fire slike tilfeller. For ytterligere to tilfeller, se Bollen (2012). I denne og påfølgende figurer vil jeg benytte bokstaven T for variabelen vi ønsker å finne den kausale effekten av, Y for utfallet og X og U for bakenforliggende variabler som er henholdsvis observerte og uobserverte. Videre benytter jeg en stjerne for å indikere at variabler inkluderer målefeil (jf. VanderWeele og Hernán 2012).



Figur 1: Fire utfordringer for kausal identifisering.

I figur 1a er problemet en uobservert konfunderende variabel. Tenk at T er hvorvidt man tar høyere utdanning, Y er lønn og U er (uobserverte) kognitive ferdigheter når man er 18 år. Ettersom kognitive ferdigheter påvirker både hvorvidt man tar høyere utdanning og lønn, sier vi at den bivariate sammenhengen mellom høyere utdanning og lønn er delvis konfundert. Den bivariate sammenhengen mellom høyere utdanning og lønn vil fange opp både den kausale sammenhengen $T \rightarrow Y$ og den ikke-kausale sammenhengen $T \leftarrow U_1 \rightarrow Y$. Mest sannsynlig vil vi overdrive styrken på sammenhengen mellom høyere utdanning og lønn, ettersom alle sammenhengene i figur 1a trolig er positive.

Hvis vi derimot observerer kognitive ferdigheter, kan vi bruke kontrollvariabelmetoden for å unngå dette problemet. Ved å kontrollere for kognitive ferdigheter blokkerer vi bakdørsveien mellom høyere utdanning og lønn som går gjennom kognitive ferdigheter, og vi står igjen med den kausale effekten av å ta høyere utdanning på lønn. I praksis har vi imidlertid imperfekte mål på de konfunderende variablene. Dette er vist i figur 1b, hvor x_1^* er et imperfekt mål på den konfunderende uobserverte variabelen U_1 . Vi kan tenke oss at U_1 er kognitive ferdigheter mens x_1^* er karakterer fra videregående utdanning. Personer med høye karakterer fra videregående utdanning har i snitt høyere kognitive ferdigheter. Dermed vil det å kontrollere for karakterer fra videregående utdanning fjerne noe av konfunderingen. Ettersom karakterer er et imperfekt mål på kognitive ferdigheter vil det imidlertid ikke være tilstrekkelig å kontrollere for karakterer for å unngå all konfundering i figur 1b (Pearl 2010). Etter å ha kontrollert for karakterer vil det fortsatt være en restkonfundering som skyldes målefeil i den konfunderende variabelen.

Målefeil i forklaringsvariabelen er et annet vanlig problem i samfunnsforskning, og er vist i figur 1c. Selv om målefeilen ikke er systematisk vil vi kunne få feilaktige koeffisienter, gitt at målefeilen er ukorrelert med den uobserverte forklaringsvariabelen T . Generelt vil usystematisk målefeil i forklaringsvariabelen føre til at estimatene tilnærmer seg null (Wooldridge 2006: 318–320).³ Ved å bruke kontrollvariabelmetoden vil vi derfor undervurdere sammenhengen mellom variablene. Dette er en bekymring blant annet i litteraturen om den økonomiske avkastningen av utdanning, og spesielt blant studier som benytter familiefasteffekter (se Card 1999: 1815–1816). Ved å benytte kontrollvariabelmetoden vil undervurderingen av estimatene *forsterkes* av å inkludere kontrollvariabler som forklare mer av variasjonene i forklaringsvariabelen (Frisell et al. 2012). Når vi benytter kontrollvariabelmetoden vil vi

derfor måtte vurdere feilkilden som følge av konfundering opp mot feilkilden som følge av målefeil, ettersom visse strategier for å unngå konfundering kan forverre konsekvensene av målefeilen.

Den fjerde og siste figuren i figur 1 viser et tilfelle av resiprok kausalitet. Vi kan i figur 1d tenke på T som medstudenters karakternivå og Y som individets karakternivå, målt på tidspunkt 1 og tidspunkt 2. Å gå i en klasse med flinke medstudenter har trolig en positiv innvirkning på eget karakternivå ($T_{t=1} \rightarrow Y_{t=2}$), men samtidig vil individets eget karakternivå påvirke medstudenters karakternivå ($Y_{t=1} \rightarrow T_{t=2}$). Hvis vi observerer T og Y samtidig er kontrollvariabelmetoden utilstrekkelig for å identifisere den kausale effekten av T på Y .

I alle disse fire tilfellene er det grunnleggende problemet det samme: restleddet i en lineær regresjonsmodell er korrelert med forklaringsvariabelen, og kontrollvariabelmetoden vil derfor gi et feilaktig bilde av sammenhengen mellom variablene (Bollen 2012). I de fleste tilfeller av samfunnsforskning vil én eller flere av disse problemene påvirke resultatene, og behovet for å benytte alternative metoder for å identifisere kausale effekter er derfor stort. Instrumentvariabelmetoden kan i teorien benyttes for å unngå alle problemer som skyldes at restleddet er korrelert med forklaringsvariabelen, og er derfor et nyttig verktøy for å identifisere kausaleffekter. I den følgende delen presenterer jeg instrumentvariabelmetoden. For å forenkle presentasjonen benytter jeg tilfellet i figur 1a som eksempel, men logikken er den samme selv om utfordringen skulle være målefeil eller resiprok kausalitet.

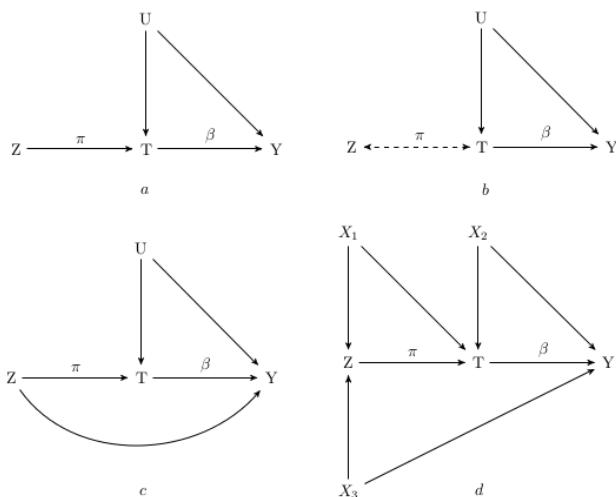
Hva er en instrumentvariabel?

Vi kan tenke oss at det er to kilder til variasjon i forklaringsvariabelen: variasjon som skyldes konfunderende variabler og variasjon som skyldes andre variabler. Når vi er opptatt av effekten av en forklaringsvariabel på en avhengig variabel, vil den siste typen av variasjon kunne tolkes som (kvasi)tilfeldig variasjon. Variabler som bidrar til denne typen variasjon omtales som eksogene variabler. Kontrollvariabelmetoden går ut på å kontrollere for alle konfunderende variabler slik at vi står igjen med en forklaringsvariabel hvor den resterende variasjonen antas å være tilfeldig.⁴ Instrumentvariabelmetoden tilnærmer seg imidlertid problemet med konfunderende variabler på en helt annen måte. Istedenfor å kontrollere for konfunderende variabler, tar denne metoden utgangspunkt i én

eller flere eksogene variablene som bidrar til den tilfeldige variasjonen i forklaringsvariabelen. Ved hjelp av eksogene variabler kan vi isolere den delen av variasjonen i forklaringsvariabelen som er upåvirket av konfunderende variabler.

Kort fortalt er en instrumentvariabel eller et instrument (Z) en variabel som er korrelert med den avhengige variabelen kun gjennom forklaringsvariabelen. Dette innebærer to forutsetninger. For det første må Z være korrelert med T . Det vil si at Z må være *relevant*. Det er mange måter Z kan være korrelert med T på. Figur 2a og 2b viser to ulike måter: Z kan ha en kausal effekt på T og Z og T kan være korrelert på grunn av bakenforliggende variabler. For det andre innebærer dette at Z verken har en direkte effekt på Y eller er korrelert med andre utelatte variabler som påvirker Y . Dette omtales som *instrumentvaliditet*, og er oppfylt i både figur 2a og 2b. Både instrumentrelevans og instrumentvaliditet gjelder kontrollert for andre variabler i modellen. Det vil si at etter å ha kontrollert for andre nødvendige variabler, må det fortsatt være en effekt av Z på T , og kun én effekt av Z på Y gjennom T . Jeg kommer tilbake til bruken av kontrollvariabler senere.

Hvis vi har et instrument som er både relevant og valid, kan det benyttes for å identifisere den kausale effekten av T på Y . For å forstå dette er det nødvendig å se nærmere på sammenhengen mellom Z og Y . La oss ta utgangspunkt i figur 2a, hvor Z er instrumentvariabelen, T er forklaringsvariabelen,



Figur 2: Instrumentvariabelmetoden.

Y er utfallet, π er effekten av Z på T og β er effekten av T på Y . Kovariansen mellom Z og Y består generelt i

$$[1] \quad \text{cov}(Z, Y) = \beta * \text{cov}(Z, T) + \text{cov}(Z, \epsilon)$$

, hvor ϵ er alle uobserverte variabler som påvirker Y . I figur 2a er Z korrelert med Y kun gjennom T ($\text{cov}(Z, \epsilon) = 0$), slik at [1] forenkles til:

$$[2] \quad \text{cov}(Z, Y) = \beta * \text{cov}(Z, T)$$

Det vil si at kovariansen mellom Z og T består av den kausale effekten av T på Y , dvs. β , ganger kovariansen mellom Z og T . Altså vil ikke sammenhengen mellom Z og Y påvirkes av konfunderende variabler som U . Dette er enklest å forstå ved å ta utgangspunkt i prinsippene bak kausale grafer. Ettersom det går en pil fra både Z og U til T , er T en kolliderende variabel. Sammenhengen mellom to variabler kan ikke gå gjennom kolliderende variabler.

Legg merke til at sammenhengen mellom Z og Y ikke er lik sammenhengen mellom T og Y . Logikken bak bruken av instrumentvariabler er at sammenhengen mellom Z og Y kan benyttes for å isolere sammenhengen mellom T og Y . Ved å dele kovariansen mellom Z og Y på kovariansen mellom Z og T oppnår vi nettopp dette:

$$[3] \quad \frac{\text{cov}(Z, Y)}{\text{cov}(Z, T)} = \frac{\beta * \text{cov}(Z, T)}{\text{cov}(Z, T)} = \beta$$

Ved å identifisere effekten av T på Y på denne måten benytter vi kun variasjon i forklaringsvariabelen T som skyldes det eksogene instrumentet Z , mens den resterende variasjonen i T ikke benyttes. Det innebærer at vi dropper all variasjon i T som er ukorrelert med Z , og dermed all variasjon i T som skyldes konfunderende variabler (Angrist og Krueger 2001: 73).⁵ Dette kan også sees ved å benytte prinsipper for stianalyse. Hvis variablene er z-standardiserte, vil kovariansen mellom to variabler kunne tolkes som korrelasjon (Brown 2006: 61). Kovariansen mellom Z og Y vil da være lik π ganger β , og ved å benytte samme logikk som i [3] ser vi at: $\frac{\pi\beta}{\pi} = \beta$.

La meg gjenta logikken bak instrumentvariabler ved å vise den vanligste metoden for å estimere IV-koeffisienter: To-steps-minste-kvadraters-metode

(Two-stage least-squares, 2SLS). Vi fortsetter med eksempelet med instrumentet Z , forklaringsvariabelen T og utfallet Y , hvor π er effekten av Z på T og β er effekten av T på Y . Når vi benytter 2SLS starter vi med en regresjonsanalyse med T som avhengig variabel. Dette omtales ofte som første-steget. I [4] er ϖ_0 konstantleddet og τ restleddet.

$$[4] \quad T = \varpi_0 + \pi Z + \tau$$

I det andre steget bruker vi de predikerte verdiene $\hat{T}$ fra denne regresjonsanalysen istedenfor T i analysen med Y som avhengig variabel:

$$[5] \quad Y = \beta_0 + \beta \hat{T} + \varepsilon$$

Tenk at vi skal undersøke den kausale effekten av å ta høyere utdanning på lønn, men ikke har informasjon om kognitive ferdigheter før observasjons-enhetene starter høyere utdanning. Ettersom personer med høye kognitive ferdigheter forventes både å ha høyere sannsynlighet for å ta høyere utdanning *og* høyere sannsynlighet for å få høy lønn senere, er resultatene konfundert. Hvorvidt en person tar høyere utdanning eller ikke T er delvis bestemt av kognitive ferdigheter U . Flere forskere har foreslått at avstand til høyere utdanningsinstitusjoner kan benyttes som et instrument for om personer tar høyere utdanning (Card 1999). Argumentet er at kostnadene ved å ta en høyere utdanning er mindre for personer som bor nær slike institusjoner, og at det således er en sammenheng mellom nærhet til utdanningsinstitusjoner Z og hvorvidt man tar høyere utdanning T . Ved at vi predikere utdanningsnivå som en funksjon av avstand får vi en variabel $\hat{T}$ som ikke er en funksjon av konfunderende variabler U . Variasjonene i de predikerte verdiene $\hat{T}$ fra [4] er kun et resultat av Z , og dermed ukorrelert med konfunderende variabler. Mens verdiene på T er delvis bestemt av konfunderende variabler, er derimot de predikerte verdiene $\hat{T}$ kun påvirket av avstand til utdanningsinstitusjoner. Ved hjelp av 2SLS kan vi dermed estimere effekten av en *kvasi-randomisert* T -variabel på Y , på tross av uobserverte konfunderende variabler.

Det er både likheter og forskjeller mellom instrumentvariabelmetoden og kontrollvariabelmetoden. På den ene siden er målet i begge metodene å stå igjen med en forklaringsvariabel hvor variasjonen ikke skyldes konfunderende variabler. På den andre siden oppnås dette på en annen måte i instrumentvariabelmetoden enn kontrollvariabelmetoden; kontrollvariabelmetoden iso-

lerer den delen av variasjonen i T som ikke er korrelert med observerte kontrollvariabler, mens instrumentvariabelmetoden isolerer den delen av variasjonen i T som er korrelert med instrumentvariabelen. Fordelen med å bruke instrumentvariabler heller enn kontrollvariabelmetoden er at vi kan unngå også andre utfordringer, som målefeil og resiprok kausalitet, i tillegg til at instrumentvariabelmetoden kan brukes når de konfunderende variablene er uobserverte. Utfordringen med å bruke instrumentvariabelmetoden er at brudd på forutsetningene kan føre til store skjevheter i estimatene. I en del tilfeller vil feilkilden ved å bruke instrumentvariabelmetoden være større enn ved å bruke kontrollvariabelmetoden. I det følgende diskuterer jeg noen sentrale utfordringer og fallgruver når man benytter instrumentvariabler.

Invalide instrumenter

For at et instrument skal være valid, må instrumentet kun påvirke Y gjennom T . Rent intuitivt er dette nødvendig fordi identifiseringen av kausale effekter i instrumentvariabelmetoden bygger på antagelsen om at *hele* sammenhengen mellom Z og Y går gjennom T . Hvis noe av sammenhengen går utenom T vil vi dermed trekke feil slutninger om sammenhengen mellom T og Y . Figur 2c gir et eksempel på et tilfelle hvor Z er invalid; det er både en effekt av Z på Y gjennom T og en direkte effekt av Z på Y . For å se på konsekvensene av dette, tar vi utgangspunkt i [1] og [3] over. Hvis deler av sammenhengen mellom Z og Y går utenom T , innebærer dette at kovariansen mellom Z og restleddet ϵ ikke er lik 0. Da vil instrumentvariabelmetoden identifisere det følgende:

$$[6] \quad \hat{\beta}_{IV} = \frac{cov(Z,Y)}{cov(Z,T)} = \frac{\beta * cov(Z,T) + cov(Z,\epsilon)}{cov(Z,T)} = \beta + \frac{cov(Z,\epsilon)}{cov(Z,T)}$$

Vi vil dermed *ikke* identifisere den kausale effekten av T på Y (β). For å tydeliggjøre dette er det fordelaktig å ta utgangspunkt i et konkret eksempel. La oss bygge videre på eksempelet om høyere utdanning og lønn, hvor avstand til utdanningsinstitusjoner benyttes som instrument. Problemet med dette instrumentet er at det å bo nærme utdanningsinstitusjoner kan ha en separat effekt på lønn som *ikke* går gjennom det å ta høyere utdanning. For eksempel kan det være at kvaliteten på grunn- og videregående skole er bedre i områder nært til høyere utdanningsinstitusjoner, og at dette påvirker lønnsnivået. Den sammenhengen vi observerer mellom Z og Y vil dermed ikke bare skyldes ef-

fekten som går gjennom T . Dette problemet kan løses hvis vi har informasjon om kvaliteten på grunnskoler og videregående skoler, ettersom vi da kan blokkere den direkte effekten av Z på Y ved å kontrollere for kvaliteten på disse skolene. Dette viser hvordan man ofte vil trenge en rekke kontrollvariabler også når man benytter instrumentvariabler.

Det er foreslått flere måter å teste instrumentvaliditet på. For det første har det blitt hevdet at man kan teste validiteten ved å undersøke om det er en signifikant effekt av Z på Y kontrollert for T . Problemet med dette er at vi ved å kontrollere for T kan skape en fiktiv sammenheng mellom Z og Y som går gjennom alle uobserverte variabler som påvirker T , ettersom T er en kolliderende variabel. For en utdypelse av dette poenget, se Morgan og Winship (2007: 196–197). En annen foreslått mulighet for å undersøke instrumentvaliditet er å ha flere instrumenter enn forklaringsvariabler. Ved å benytte en test av overidentifisering kan vi undersøke om vi oppnår samme parameterestimat hvis vi benytter to eller flere forskjellige instrumenter. Problemet med denne metoden er at to forskjellige instrumenter vil identifisere det samme parameterestimatet kun under visse forutsetninger, slik som effekthomogenitet. Igjen henviser jeg til Morgan og Winship (2007) for en grundigere diskusjon. Forutsetningen om effekthomogenitet utdypes ytterligere under.

Forutsetningen om instrumentvaliditet lar seg dermed ikke teste på noen tilfredsstillende måte. I all hovedsak må vi dermed begrunne instrumentvaliditeten på bakgrunn av teori og tidligere forskning. Dette krever at vi har en god forståelse av forskningsfeltet. Det er ikke unikt for instrumentvariabelmetoden at en viktig forutsetning for å identifisere den kausale effekten ikke lar seg teste. All kausal identifisering bygger på en eller flere forutsetninger (Pearl 2009). For eksempel forutsetter man i kontrollvariabelmetoden at man evner å blokkere alle bakdørsveier mellom T og Y . Problemet med instrumentvariabelmetoden er at skjevheten i parameterestimatene kan bli langt større enn i kontrollvariabelmetoden hvis instrumentet er invalid, spesielt hvis instrumentet er svakt (Cameron og Trivedi 2005: 106). Dette diskuteres i de to neste delene.

Svake instrumenter

For å kunne identifisere den kausale effekten ved hjelp av en instrumentvariabel er det avgjørende at instrumentet er korrelert med forklarings-

variabelen. Hvis instrumentvariabelen ikke er korrelert med forklaringsvariabelen er instrumentet irrelevant og det er ikke mulig å identifisere den kausale effekten ved hjelp av instrumentet. Vi kan se dette med utgangspunkt i [3] over. Hvis $cov(Z, T) = 0$ vil både teller og nevner bli 0 i [3], og vi kan ikke identifisere β . Det er kanskje enda enklere å se dette ved å ta utgangspunkt i det første steget [4] i 2SLS. Er det ingen sammenheng mellom Z og T vil det heller ikke være noe variasjon i de predikerte verdiene av T . Det vil si at alle observasjonene i populasjonen har samme verdi på $\hat{T}$. Ettersom man ikke kan identifisere effekten av en konstant (Firebaugh 2008: 37), vil vi dermed ikke kunne bruke instrumentvariabelmetoden for å identifisere effekten av T på Y hvis instrumentet er irrelevant.

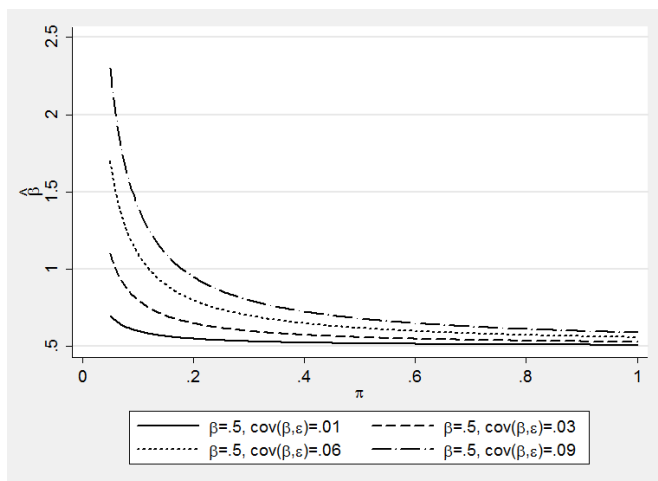
I motsetning til instrumentvaliditet kan instrumentrelevans testes. Den enkleste måten å teste instrumentrelevans på er å gjennomføre en T-test av effekten av Z på T kontrollert for de andre kontrollvariablene. Det vil si at vi tester om koeffisienten π i [4] er statistisk signifikant. At det er en signifikant sammenheng mellom Z og T er en nødvendig for å identifisere β , men ikke tilstrekkelig. Instrumentet bør også være relativt sterkt. Det er to grunner til dette. For det første benytter instrumentvariabelmetoden kun den delen av variansen i T som skyldes det eksogene instrumentet Z . Ettersom instrumentvariabelmetoden bruker en mindre andel av variansen, vil også standardfeilen bli større enn ved å bruke kontrollvariabelmetoden (Morgan og Winship 2007: 198). Mens resultater som bruker kontrollvariabelmetoden vil gi inkonsistente men presise resultater, vil instrumentvariabelmetoden gi konsistente men upresise parameterestimater (gitt at instrumentet er valid). Med mindre utvalgsstørrelsen er stor og/eller den faktiske effekten er sterk, vil risikoen for å gjøre feil av type II dermed være stor når vi bruker instrumentvariabler. Med andre ord vil vi ha større sannsynlighet for feilaktig å konkludere med at det ikke er noen sammenheng mellom T og Y når vi benytter instrumentvariabelmetoden enn når vi benytter kontrollvariabelmetoden. Poenget er at hvis instrumentet er svakt nok vil ikke estimatene være særlig informative (Murray 2006: 115). Jo sterkere instrumentet er, jo mer av variasjonene i forklaringsvariabelen benyttes, og desto mindre blir risikoen for feil av type II.

Svake og invalide instrumenter

For det andre avhenger konsekvensene av invalide instrumenter av instrumen-

tets styrke. For å tydeliggjøre betydningen av dette, se igjen på [6] som bygger på figur 2c hvor instrumentet er invalid. Som vi allerede har sett, vil vi ved å benytte instrumentet Z identifisere den kausale effekten av T på Y (β) og en feilkilde som skyldes den ikke-kausale effekten som går direkte fra Z til Y ($\frac{\text{cov}(Z, \epsilon)}{\text{cov}(Z, T)}$). La oss nå se nærmere på hva som påvirker størrelsen på feilkilden.

Figur 3 viser hvordan feilkilden påvirkes av både instrumentets validitet og instrumentets styrke. Y-aksen viser $\hat{\beta}$, som er den estimerte effekten av T på Y ved å bruke instrumentvariabelmetoden, mens X-aksen viser styrken på instrumentet π . Effekten av forklaringsvariabelen T er holdt konstant på .5 i figuren, mens korrelasjonen mellom Z og restleddet ($\text{cov}(Z, \epsilon)$) varierer fra .01 til .09. Ut av figuren kan vi lese to ting. For det første er feilkilden større når korrelasjonen mellom Z og ϵ er sterkest, som vi forventer. For det andre ser vi at feilkilden er svært sensitiv til styrken på Z , særlig når det er en sterk korrelasjon mellom Z og ϵ . Med svake instrumenter vil feilkilden kunne bli svært stor, selv om korrelasjonen mellom instrumentet og restleddet er liten. Dette problemet påvirkes *ikke* av størrelsen på datasettet (Morgan og Winship 2007: 199).



Figur 3: Skjevhet i instrumentvariabelestimateret $\hat{\beta}$ som følge av brudd på forutsetningen om instrumentvaliditet ($\text{cov}(\beta, \epsilon) = 0$) etter styrken på instrumentet π .

Instrumentvariabelmetoden benyttes primært fordi man har gode grunner til å forvente at kontrollvariabelmetoden vil gi feilaktig estimater. Det kan derfor være nyttig å sammenligne feilkilden i instrumentvariabelmetoden ([6]), med den tilsvarende feilkilden i kontrollvariabelmetoden:

$$[7] \quad \hat{\beta}_{OLS} = \beta + \frac{cov(T, \epsilon)}{var(T)}$$

Ved å sammenligne [6] og [7] ser vi at hvis instrumentet er svakt, vil selv en svak sammenheng mellom Z og ϵ kunne føre til at feilkilden er større i instrumentvariabelmetoden enn i kontrollvariabelmetoden. Wooldridge (2006: 514–515) gir et eksempel på dette. Hvis korrelasjonen mellom instrumentet og forklaringsvariabelen er .2 må korrelasjonen mellom instrumentet og restleddet være mindre enn en femtedel av korrelasjonen mellom forklaringsvariabelen og instrumentet for at instrumentvariabelmetoden skal ha mindre feilkilde enn kontrollvariabelmetoden. Ofte benyttes instrumenter som har en langt svakere korrelasjon med forklaringsvariabelen enn .2 (se eksempel i Bound et al. 1995). Dette krever at korrelasjonen mellom instrumentet og restleddet må være enda mindre relativt til korrelasjonen mellom forklaringsvariabelen og restleddet. Hvis korrelasjonene går i samme retning må generelt $cor(Z, \epsilon)/cor(Z, T) < cor(T, \epsilon)$ for at feilkilden skal være mindre i instrumentvariabelmetoden enn i kontrollvariabelmetoden. Merk at ettersom vi ikke observerer feilleddet ϵ , vil vi aldri vite hvor stor feilkilden er i de to tilnærmingene.

Det er mye som taler for å være forsiktig med å benytte instrumentvariabelmetoden i alle tilfeller hvor instrumentet er svakt, selv om vi har gode grunner til å tro at instrumentet er valid. Morgan og Winship (2007: 199–199) gir en klar og tydelig begrunnelse for dette. Det er lite sannsynlig at kovariansen mellom et instrument og en forklaringsvariabel er null i endelige utvalg. Vi vil derfor alltid kunne estimere parameterne, selv om instrumentet er irrelevant. I slike tilfeller vil standardfeilen underestimeres. Videre vil små brudd på validitetsforutsetningen om at instrumentet bare skal påvirke utfallsvariabelen gjennom forklaringsvariabelen kunne føre til store skjevheter i estimatene, som vi har sett over. Er instrumentet svakt nok vil feilkilden kunne bli betydelig også med svært store utvalg, både når det gjelder punktestimater og standardfeil (Bound et al. 1995, Angrist og Pischke 2009: 205–209). Studier som benytter instrumenter fra naturlige eksperimenter (for en beskrivelse, se Angrist og Krueger 2001) er ikke immun mot disse problemene.

Randomiseringen i naturlige eksperimenter er ikke i seg selv en garanti for at instrumentet ikke har en direkte effekt på utfallsvariabelen (Morgan og Winship 2007: 199), samtidig som instrumenter fra naturlige eksperimenter ofte er svake.

I praksis vil instrumentet nesten alltid kunne være (svakt) korrelert med feilleddet, noe som innebærer at det er viktig at instrumentet er sterkt.⁶ Når instrumentet er sterkt, vil nemlig små brudd på instrumentvaliditeten ikke gi for store skjevheter i estimatene (Murray 2006: 128), noe som kan sees av figur 3. Dessverre er det ingen klare retningslinjer for å vurdere hva som regnes som et *for* svakt instrument. Et minstekrav kan være at det er en signifikant effekt av Z på T , men dette er ikke nødvendigvis tilstrekkelig for å unngå betydelig feilestimering. Er datasettet stort nok vil en svak instrumenteffekt kunne bli signifikant, og selv et lite brudd på instrumentvaliditeten vil kunne få store konsekvenser (Morgan og Winship 2007: fotnote 14 side 199). Se Cameron og Trivedi (2005) for noen forslag til hvordan man kan vurdere svake instrumenter.

Effektheterogenitet

Sammenhenger mellom variabler i samfunnsforskning er som oftest kontekstavhengig. Effekten av en variabel varierer over individer, mellom grupper og over tid og rom (Elwert og Winship 2010). Dette omtales som effektheterogenitet, og kan potensielt ha store konsekvenser for tolkningen av resultatene i instrumentvariabelmetoden. La oss anse T som en behandling. Ettersom vi arbeider med observasjonsdata, har vi ikke kontroll over hvem som skal motta behandlingen. Individer i utvalget kan puttes i fire undergrupper: de som velger en behandling på grunn av instrumentet (compliers), de som unngår å velge en behandling på grunn av instrumentet (defiers), de som vil velge en behandling uansett (always takers) og de som aldri vil velge en behandling (never takers). I eksempelet med avstand til utdanningsinstitusjoner som instrument, er compliers de som velger å ta høyere utdanning fordi de bor nær en utdanningsinstitusjon, defiers er de som lar være å ta høyere utdanning («trass») fordi de bor nær en utdanningsinstitusjon, mens always takers og never takers ikke påvirkes av instrumentet. Generelt er vi opptatt av de som velger en gitt behandling på grunn av instrumentet, nemlig compliers.⁷

Når vi benytter instrumentvariabler til å estimere kausale effekter vil kun individer som påvirkes av instrumentet bidra til estimatet. Tenk på personer hvis utdanningsvalg ikke påvirkes av om de bor nærme eller langt unna en utdanningsinstitusjon (never takers og always takers). Å vite verdien på Z for disse personene vil ikke fortelle oss noe om de har tatt høyere utdanning, ettersom utdanningslengden deres ikke påvirkes av instrumentet. De vil derfor være tilfeldig fordelt innen Z (Morgan og Winship 2007: 203). Derimot vil det å vite avstanden til utdanningsinstitusjoner for personer som påvirkes av instrumentet fortelle oss noe om personene har tatt høyere utdanning. Ettersom vi i instrumentvariabelmetoden identifiserer effekten av å ta høyere utdanning på lønn ved å benytte variasjon i utdanningslengde som følge av instrumentet, innebærer dette at kun personer som påvirkes av instrumentet bidrar til identifisering av den kausale effekten. La oss for nå også anta at alle i populasjonen som påvirkes av instrumentet er compliers. Det vi identifiserer med instrumentvariabelmetoden er en lokal gjennomsnittlige behandlingseffekt for de som blir påvirket enten positivt eller negativt av instrumentet (local average treatment effect, LATE) (Angrist og Pischke 2009: 151–161).

Når det er effektheterogenitet kan vi dermed ikke benytte instrumentvariabler for å identifisere en gjennomsnittlig kausaleffekt for alle i populasjonen, ettersom alle som er upåvirket av instrumentet (always takers og never takers) faller ut. Dette regnes ofte som en innvending mot å bruke instrumentvariabler. Angrist og Krueger (2001) argumenterer imidlertid for at man kan snu dette på hodet. I tilfeller hvor det er effektheterogenitet er det ikke nødvendigvis tilfellet at en gjennomsnittlig kausaleffekt belyser forskningsspørsmålet på en god måte. Derimot kan instrumentvariabelmetoden identifisere den ønskede kausaleffekten. La oss ta et eksempel. Angrist og Krueger (1991) estimerer effekten av utdanning på inntekt ved å bruke fødselstidspunkt som instrument. Barn starter i utdanning i det kalenderåret de fyller seks år i de fleste stater i USA, slik at de som er født sent på året er yngre når de starter på skolen enn de som er født tidlig på året. Videre er det slik at lover om obligatorisk utdanning fører til at man må fortsette i utdanning til man fyller 16 år. Altså vil de som er født sent på året tvinges til å ta lengre utdanning enn de som er født tidlig på året, og dermed ha litt lengre utdanningsnivå. Angrist og Krueger (1991) benytter derfor fødselstidspunkt som et instrument for utdanning for å undersøke effekten av utdanning på inntekt. Det som identifiseres med denne metoden er om de som tvinges til å ta litt lengre utdanning som følge av at de er født sent på året har høyere inntekt enn de som

er født tidligere på året. Denne gruppen av compliers er åpenbart ikke generaliserbar til hele populasjonen; effekten av utdanning for de som må tvinges til å ta lengre utdanning kan godt være annerledes enn effekten av utdanning for de som er villig til å ta lengre utdanning. Det er for eksempel rimelig å anta at læringsutbyttet av utdanning avhenger av innsatsen elevene legger inn i utdanningen. Instrumentet vil dermed ikke fange opp den gjennomsnittlige kausaleffekten av utdanning på inntekt for hele populasjonen. Hvis forskningsspørsmålet går ut på å undersøke konsekvensene av å øke antall år med obligatorisk utdanning vil imidlertid den gjennomsnittlige kausaleffekten for hele populasjonen være lite informativ. Den lokale kausaleffekten som identifiseres med instrumentet vil derimot fange opp effekten for den gruppen som påvirkes av lovgivningen, som er den gruppen som er interessant med tanke på forskningsspørsmålet. Dette er et eksempel på at instrumentvariabler krever faglig innsikt og kreativitet fra forskerne.

Monotonitet

Over antok vi at alle personene som påvirkes av instrumentet er compliers. Dette er monotonitetsforutsetningen (monotonicity assumption) (Imbens og Angrist 1994). Ofte vil imidlertid populasjonen bestå av både compliers og defiers. Chaisemartin og d'Haultfoeuille (2012) gir et eksempel på dette, med utgangspunkt i en studie av Angrist og Evans (1998) som undersøker om det å få barn påvirker foreldres arbeidsmarkedsdeltakelse. Angrist og Evans (1998) argumenterer for at foreldre ønsker en søskenflokk bestående av både gutter og jenter, slik at foreldre som får to barn av samme kjønn har større sannsynlighet for å få et tredje barn enn foreldre som får to barn av forskjellig kjønn. Kjønnssammensetningen på de to første barna kan derfor benyttes som et instrument. I deres studie vil dermed foreldre som velger å få et tredje barn som følge av at de to første er av samme kjønn være compliers, de som velger å få et tredje barn uavhengig av kjønnnet på de to første være always takers og de som uansett ikke vil få flere barn uavhengig av kjønnnet være never takers. Det er imidlertid også grunn til å forvente at det er en gruppe defiers i populasjonen. Foreldre som får to jenter har større sannsynlighet for å få et tredje barn enn foreldre som får to gutter, ettersom noen foreldre har en preferanse for gutter. Foreldre som er compliers vil derfor velge å få et tredje barn fordi de har fått to gutter, mens defiers vil ikke få et tredje barn hvis de allerede har fått to gutter.

Monotonitetsforutsetningen er nødvendig for å ha en veldefinert gruppe som påvirkes av instrumentet. Hvis denne forutsetningen ikke er oppfylt vet vi rett og slett ikke hvem den lokale kausaleffekten gjelder for. I tillegg er det en mer subtil grunn til at gruppen må bestå av kun compliers. For å se dette tar vi utgangspunkt i [8] og [9] hentet fra Angrist et al. (1996), hvor vi forutsetter at instrumentet er relevant og valid. Vi skal bruke [8] og [9] til å vise konsekvensene av effektheterogenitet og/eller brudd på monotonitetsforutsetningen. I [8] er $\beta_{compliers}$ den kausale effekten for compliers, $\beta_{defiers}$ den kausale effekten for defiers og λ viser andelen defiers blant de som påvirkes av instrumentet. Vi ønsker å identifisere den kausale effekten for compliers ($\beta = \beta_{compliers}$)

$$[8] \quad \hat{\beta} = (1 + \lambda) * \beta_{compliers} - \lambda * \beta_{defiers}$$

$$[9] \quad \text{hvor } \lambda = \frac{P(i \text{ er defier})}{P(i \text{ er complier}) - P(i \text{ er defier})}$$

Ut i fra [8] kan vi trekke to slutninger. For det første vil brudd på monotonitetsforutsetningen påvirke den estimerte effekten $\hat{\beta}$. Dette fanges opp av λ . Fra [9] ser vi at hvis andelen defiers er lik null, vil også λ bli lik null, og vi står igjen med $\hat{\beta} = \beta_{compliers}$ i [8]. Er derimot andelen defiers ikke lik null, vil heller ikke λ være lik null. Størrelsen på λ avhenger i dette tilfellet av andelen defiers relativt til andelen compliers. Igjen ser vi et eksempel på at det er viktig at instrumentet er sterkt. For det andre påvirkes $\hat{\beta}$ av effektheterogenitet mellom compliers og defiers. Hvis effekten av T på Y er relativt lik for compliers og defiers, vil skjevheten bli liten. Og hvis det er ingen forskjell, det vil si effekthomogenitet, vil $\hat{\beta} = \beta_{compliers}$.

Vi er med andre ord trygge hvis det enten er effekthomogenitet *eller* hvis monotonitetsforutsetningen er oppfylt. Dette er mer komplisert enn de andre forutsetningene i instrumentvariabelmetoden. En omorganisering av ligning [8] til ligning [10] kan være tydeliggjørende:

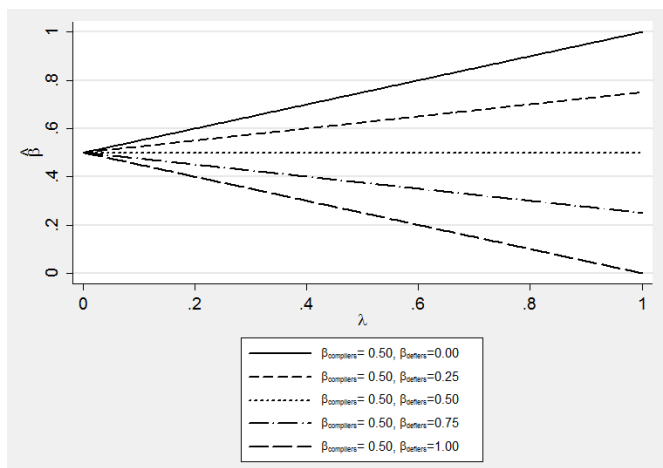
$$[10] \quad \hat{\beta} = \beta_{compliers} + \lambda (\beta_{compliers} - \beta_{defiers})$$

Den siste delen (feilleddet) faller bort hvis enten $\lambda=0$ *eller* $\beta_{compliers} - \beta_{defiers}=0$. Altså, på den ene siden har effektheterogenitet ingen betydning hvis vi kun

bruker variasjonen som skyldes compliers ($\lambda=0$), mens på den andre siden har det ingen betydning at populasjonen består av både compliers og defiers hvis den kausale effekten er lik for disse gruppene ($\beta_{compliers} - \beta_{defiers} = 0$). Hvis det derimot er effektheterogenitet *og* populasjonen består av både compliers og defiers, noe som er sannsynlig, vil vi få skjeve estimater.

Retningen på skjevheten er kontraintuitiv når monotonitetsforutsetningen ikke er oppfylt og det er effektheterogenitet. Derfor er det viktig at denne forutsetningen forstås godt. Det naturlige ville være å forvente at vi får et enkelt gjennomsnitt av de to gruppenes kausaleffekt. Dette er imidlertid *ikke* tilfellet. Faktisk er det mulig at effekten er positiv for både compliers og defiers, men den estimerte effekten blir negativ ("sign reversal"). La meg kort eksemplifisere dette. I figur 4 tar jeg utgangspunkt i [8] og holder verdien for $\beta_{compliers}$ konstant på .5, mens jeg endrer verdien for λ og $\beta_{defiers}$. Verdiene på Y-aksen viser den estimerte effekten av β , mens verdiene på X-aksen viser endringer i λ . La oss først ta utgangspunkt i den heltrukne linjen, hvor vi forutsetter at effekten for compliers og defiers er henholdsvis .5 og .0. I motsetning til hva vi vil forvente går *ikke* den estimerte effekten $\hat{\beta}$ mot 0 når antallet defiers øker. Tvert om, effekten går mot 1. Det motsatte er tilfellet når vi setter kausaleffekten for defiers til 1; da går effekten mot 0.⁸ Dette innebærer at vi overdriver β hvis effekten er *svakere* for defiers enn compliers, mens vi undervuderer β hvis effekten er *sterkere* for defiers enn compliers. Feilkilden er størst når antall defiers er lik antall compliers, og feilkilden avhenger av nivået av effektheterogenitet.

Vi kan trekke to implikasjoner ut av det vi har sett her. Ettersom vi sjeldent kan forvente effekthomogenitet er det viktig at monotonitetsforutsetningen er oppfylt. Hvis forutsetningen om monotonitet brytes sitter vi igjen med resultater uten noen veldefinert kausal fortolkning, og selv skjeve resultater fra kontrollvariabelmetoden kan være mer informative. Den andre implikasjonen er mer positiv, og innebærer at vi kan benytte informasjonen i [8] til å vurdere resultatene. Vi ser for eksempel at hvis vi forventer at effekten for defiers er dobbelt så stor som effekten for compliers, vil λ måtte være større enn 1 for å snu fortegnet på sammenhengen. Eller motsatt, hvis λ er lik 1 må effekten for defiers være minst dobbelt så stor som effekten for compliers for å snu fortegnet. Angrist et al. (1996) illustrerer med utgangspunkt i en tidligere publisert studie hvordan denne logikken kan brukes for å vurdere resultatene fra instrumentestimering.



Figur 4: Skjevhet i instrumentvariablestimater $\hat{\beta}$ som følge av brudd på monotonitetsforutsetningen og/eller effektheterogenitet.

Flere instrumenter enn endogene variabler

Jeg har til nå forutsatt at vi har hatt ett instrument og én endogen variabel. Hvis det er like mange instrumenter som endogene variabler sier man i litteraturen at modellen er akkurat-identifisert ("just-identified"). Det er imidlertid ingenting som hindrer oss i fra å inkludere flere instrumenter for T i instrumentvariabelmetoden. Med flere instrumenter enn endogene variabler sier man at modellen er overidentifisert ("overidentified"). Det vi oppnår med en overidentifisert modell er at vi forklarer mer av variasjonene i T , og dermed får mer presise estimater. Dette er viktig, ettersom en svakhet med instrumentvariabelmetoden er upresise estimater. Dessverre vil overidentifisering kunne ha negative konsekvenser. For det første vil svake instrumenter forsterke problemene diskutert i delen om svake og invalide instrumenter over (Bound et al. 1995, Angrist og Pischke 2009: 205–209). Bollen (2012) anbefaler at man er spesielt forsiktig med å bruke mange instrumenter når utvalget er lite ($N \leq 100$) og når hvert ekstra instrument har et lite selvstendig bidrag til å forklare variasjonene i T . Samtidig anerkjenner han at tidligere forskning kun gir støtte til tentative anbefalinger. Det kan også være lurt å benytte andre

estimatorer enn 2SLS når man har mange svake instrumenter, som for eksempel LIML (Angrist og Pischke 2009: 209–216) eller Fuller's estimator (Murray 2010).

Behovet for å bruke mange instrumenter er størst i de samme tilfellene hvor de negative konsekvensene potensielt er størst; når utvalget er lite og instrumentene er svake. Derimot når utvalget er stort og instrumentene er sterke, er de negative konsekvensene mindre, samtidig som det er et mindre behov for å bruke mange instrumenter. Dette taler mot å bruke flere instrumenter enn endogene variabler. En ytterligere grunn til å være forsiktig med å bruke flere instrumenter enn forklaringsvariabler er problemer knyttet til effekt-heterogenitet. Forskjellige instrumenter vil kunne identifisere ulike lokale kausaleffekter. Hvis vi benytter to eller flere instrumenter må det derfor være effekthomogenitet for at vi skal unngå å ende opp med parameterestimerer som er en blanding av ulike LATE (Morgan og Winship 2007: 212). I sum er det derfor gode grunner til å være forsiktig med å benytte flere instrumenter enn endogene variabler.

Kontrollvariabler i instrumentvariabelmetoden

Stort sett har jeg forutsatt at vi ikke trenger kontrollvariabler i gjennomgangen over. I praksis har imidlertid kontrollvariabler en sentral plass i instrumentvariabelmetoden. I kontrollvariabelmetoden benyttes kontrollvariabler for å unngå bakenforliggende variabler som påvirker både T og Y , mens i instrumentvariabelmetoden benyttes kontrollvariabler for å unngå at Z og Y er korrelert på andre måter enn gjennom T (Angrist og Pischke 2009: 175). Ofte vil det derfor være forskjellige kontrollvariabler som bør inkluderes i instrumentvariabelmetoden og i kontrollvariabelmetoden. La oss ta utgangspunkt i figur 2d for å tydeliggjøre dette. Denne figuren inneholder tre potensielle kontrollvariabler (X_{1-3}). Variabelen X_3 må vi kontrollere for hvis vi skal benytte instrumentvariabelmetoden, hvis ikke kan vi få svært skjeve resultater. De to interessante variablene i denne figuren er X_1 og X_2 . For å identifisere den kausale effekten av T på Y med kontrollvariabelmetoden er det i denne modellen nødvendig å kontrollere for X_2 . I instrumentvariabelmetoden er det derimot ikke nødvendig å kontrollere for X_2 for å identifisere kausaleffekten. X_1 er et eksempel på en variabel som vi ikke bør

kontrollere for hvis vi skal benytte instrumentvariabelmetoden, men som er ubetydelig om vi inkluderer i modellen hvis vi benytter kontrollvariabelmetoden. Hvis vi kontrollerer for X_1 kontrollerer vi bort noe av sammenhengen mellom Z og T , med de følgende svake instrumenter forårsaker. Dette viser at man bør ha en klar idé om hvilke variabler som inkluderes i modellen når man benytter seg av instrumentvariabler, akkurat som man bør ha i kontrollvariabelmetoden.

Selv om vi ikke trenger å kontrollere for X_2 når vi bruker instrumentvariabelmetoden, kan det likevel være en gevinst ved å gjøre det. En utfordring ved å bruke instrumentvariabler er at vi kun benytter den delen av variasjonen i T som skyldes Z , som fører til at vi får et upresist parameterestimat for T . Når vi beregner standardfeilen til dette parameterestimatet for T benyttes imidlertid den ikke-instrumenterte T variabelen for å konstruere residualene, ikke de predikerte T verdiene (Angrist og Pischke 2009: 140). Dermed vil vi få mer presise estimater ved å kontrollere for X_2 . Ettersom de predikerte verdiene av T ikke benyttes for å estimere standardfeilen til T , vil standardfeilen bli gal hvis vi i en regresjonsanalyse bruker de predikerte verdiene av T direkte. Selv om dette kan korrigeres er det anbefalt å benytte et statistikkprogram som kan gjennomføre 2SLS for å unngå feil.

Når vi inkluderer kontrollvariabler i 2SLS er det viktig at disse inkluderes i både det første og det andre steget. Hvis ikke vil estimatene i det andre steget mest sannsynlig bli skjeve (Angrist og Pischke 2009: 189–190). Dette innebærer at instrumentet må være relevant og valid også etter inkludering av kontrollvariabler. Det er derfor viktig å ta høyde for potensielle kontrollvariabler i vurderingen av instrumentets relevans og validitet. En positiv implikasjon av dette er at vi faktisk kan "skape" et instrument ved å kontrollere for riktig sett av variabler. Bollen (2012) skiller mellom det han kaller auxiliære instrumentvariabler, instrumentvariabler fra naturlige eller randomiserte eksperimenter og modellbaserte instrumentvariabler. Prinsippene bak modellbaserte instrumentvariabler er nært beslektet med prinsippene bak kausale grafer (Pearl 2009), og går ut på å starte med å skissere de forventede sammenhengene mellom variabler i en modell. Basert på disse sammenhengene kan vi identifisere variabler som kan benyttes som instrumenter, og eventuelt hvilke kontrollvariabler som er nødvendig for å oppnå validitet og relevans. Se for eksempel Chan og Kuroki (2010) for kreativ identifisering av instrumenter ved hjelp av kausale grafer. I vurderingen er det

viktig å ta høyde for om potensielle kontrollvariabler er eksogene. I utgangspunktet bør vi ikke kontrollere for en endogen variabel, særlig hvis variabelen påvirkes direkte av instrumentet. Se Frölich (2008) for en diskusjon av endogene kontrollvariabler i instrumentvariabelmetoden.

Konklusjon

Samfunnsforskning handler i stor utstrekning om årsakssammenhenger og det er derfor viktig å benytte metoder som gjør det mulig å avdekke slike sammenhenger. Vi har nå sett at problemer som skyldes korrelasjon mellom forklaringsvariabelen og restleddet kan unngås hvis vi har en instrumentvariabel som oppfyller de følgende to forutsetningene:

- (1) Instrumentrelevans: Z har en effekt på T gitt kontrollvariablene.
- (2) Instrumentvaliditet: Z har kun en effekt på Y gjennom T gitt kontrollvariablene.

I tilfeller med effektheterogenitet må i tillegg en tredje forutsetning være oppfylt for at parameterestimaten skal ha en veldefinert kausal fortolkning:

- (3) Monotonitet: Z har enten kun en positiv effekt på T eller en negativ effekt på T .

Hvis disse forutsetningene er oppfylt kan vi benytte instrumentvariabelmetoden for å identifisere en lokal kausaleffekt, på tross av at konfunderende variabler er (delvis) uobserverte, det er målefeil i forklaringsvariabelen og/eller det er tilfeller av resiprok kausalitet. Instrumentvariabelmetoden er derfor et nyttig verktøy for å identifisere kausale effekter. Hvis det er effektheterogenitet i populasjonen er det i tillegg nødvendig å vurdere hvorvidt instrumentet identifiserer en kausaleffekt som er interessant for å belyse forskningsspørsmålet.

All kausal identifisering bygger på kvalitative forutsetninger som man ikke kan vite med sikkerhet at er oppfylt (Pearl 2009). Når man sammenligner ulike identifikasjonsstrategier er det derfor viktig å vurdere hvilke strategier som har de mest realistiske forutsetningene, men også hva som er konsekvensene av brudd på disse forutsetningene. Hovedproblemet med å benytte

instrumentvariabelmetoden er at små brudd på forutsetningene kan føre til at man får svært skjeve resultater. Selv om det er sjeldent at forutsetningene i kontrollvariabelmetoden er oppfylt, kan derfor feilkildene i kontrollvariabelmetoden likevel være langt mindre enn i instrumentvariabelmetoden. Den største utfordringen ved å benytte instrumentvariabler er å finne instrumenter som oppfyller (1) og (2), samtidig som man med relativ stor sikkerhet kan vurdere (3). Dette krever at forskeren er kreativ og har en god forståelse av forskningsområdet (Angrist og Pischke 2009: 177). Ettersom forutsetning (2) og (3) ikke lar seg teste på noen tilfredsstillende måte er forskerens kunnskap om forskningsområdet avgjørende for vellykket bruk av instrumentvariabler. Det er med andre ord ikke de statistiske eller tekniske utfordringene som er størst når man benytter instrumentvariabler, men de teoretiske.

Noter

1. Takk til Solveig T. Borgen, til tidsskriftets to anonyme konsulenter, til redaktørene og til Ferdinand Andreas Mohn for verdifulle innspill.
2. Angrist og Pischke (2009: 174–204) anbefaler å benytte lineær instrumentvariabelregresjon også med dikotome avhengige variabler. Gjennomgangen i denne artikkelen vil derfor også være relevant for forskningsopplegg med dikotome avhengige variabler.
3. Angrist og Pischke (2009) gir en intuitiv forklaring på dette. All variasjon i den korrekte målte T-variabelen T er korrelert med Y (av forutsetning). Når det gjelder den ukorrekt målte T-variabelen T^* er det imidlertid ingen grunn til å forvente at de tilfeldige målefeilene er korrelert med Y . Altså vil den delen av variasjonen som skyldes målefeil ikke være korrelert med Y , mens den korrekte målte variasjonen er korrelert med Y . Jo større målefeilen er, desto mindre andel av variasjonene i T^* er korrelert med Y . I det ekstreme tilfellet hvor T^* bare består av tilfeldige feil, vil T^* ikke ha noen effekt på Y , og dermed vil effekten være lik 0. En annen måte å se dette på er ved å benytte stianalyse. Hvis $\text{cor}(T^*, T) = 1$ (dvs. ikke målefeil), er $\text{cor}(T^*, Y) = \text{cor}(T, Y)$. Er derimot $\text{cor}(T^*, T) < 1$ (dvs. målefeil), er $\text{cor}(T^*, Y) = \text{cor}(T^*, T) * \text{cor}(T, Y)$. Dersom $\text{cor}(T^*, T) < 1$ er altså $\text{cor}(T^*, Y) < \text{cor}(T, Y)$. Dette vil si at hvis det er målefeil i T , er korrelasjonen mellom T^* og Y mindre enn korrelasjonen mellom T og Y .
4. En måte å forstå dette på er ved hjelp av Frisch-Waugh-Lovell teoremet i OLS (se for eksempel Morgan & Winship 2007: 137). Man starter med å estimere effekten av kontrollvariablene på T og Y . Dernest beregner man de predikerte residualene for T ($\tilde{T} = T - \hat{T}$) og Y ($\tilde{Y} = Y - \hat{Y}$) basert på resultatene. Til slutt estimerer man effekten av residualene for T på residualene for Y . Dette innebærer at man fjerner variasjon i T som skyldes de observerte variablene.

5. Til sammenligning estimeres koeffisienten i OLS, den mest brukte varianten av kontrollvariabelmetoden, på følgende måte når man har en uavhengig variabel:

$$\beta_{OLS} = \frac{\text{cov}(Y, T)}{\text{cov}(T, T)} = \frac{\text{cov}(Y, T)}{\text{var}(T)}.$$

6. Et unntak er randomiserte eksperimenter, men her er det heller ingen grunn til å forvente svake instrumenter (Angrist & Pischke 2009: 161).
7. Dessverre er det ikke mulig å identifisere de individuelle compliers i populasjonen (Angrist et al. 1996: 449). Det er imidlertid mulig å beregne størrelsen på gruppen av compliers og trekk ved denne gruppen. Se Angrist og Pischke (2009: 166-172) for en fremgangsmåte.
8. Figuren viser også at instrumentet identifiserer den korrekte kausaleffekten når andelen defiers er lik 0 og/eller den kausale effekten for defiers er lik den kausale effekten for compliers.

Referanser

- Angrist, Joshua D. & Evans, William N. (1998). Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size. *The American Economic Review* 88: 450–477.
- Angrist, Joshua D., Imbens, Guido W. & Rubin, Donald B. (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association* 91: 444–455.
- Angrist, Joshua D. & Krueger, Alan B. (1991). Does Compulsory School Attendance Affect Schooling and Earnings? *The Quarterly Journal of Economics* 106: 979–1014.
- Angrist, Joshua D. & Krueger, Alan B. (2001). Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *The Journal of Economic Perspectives* 15: 69–85.
- Angrist, Joshua D. & Pischke, Jörn-Steffen (2009). *Mostly harmless econometrics: an empiricist's companion*, Princeton: Princeton University Press.
- Bollen, Kenneth A. (2012). Instrumental Variables in Sociology and the Social Sciences. *Annual Review of Sociology* 38: 37–72.
- Bound, John, Jaeger, David A. & Baker, Regina M. (1995). Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogeneous Explanatory Variable is Weak. *Journal of the American Statistical Association* 90: 443–450.
- Brown, Timothy A. (2006). *Confirmatory factor analysis for applied research*, New York: Guilford Press.
- Cameron, Adrian Colin & Trivedi, Pravin K. (2005). *Microeconometrics: methods*

- and applications* Cambridge: Cambridge University Press.
- Cameron, Adrian Colin & Trivedi, Pravin K. (2010). *Microeconometrics using Stata*, College Station, Tex.: Stata Press.
- Card, David (1999). Chapter 30 The causal effect of education on earnings. I: ORLEY, C. A. & DAVID, C. (eds.) *Handbook of Labor Economics*. Elsevier.
- Chaisemartin, Clément De & D'haultfoeuille, Xavier (2012). Late Again with Defiers. *Working paper*.
- Chan, Hei & Kuroki, Manabu (2010). Using Descendants as Instrumental Variables for the Identification of Direct Causal Effects in Linear SEMs. *International Conference on Artificial Intelligence and Statistics (AISTATS)*. Chia Laguna Resort, Sardinia, Italy.
- Elwert, Felix & Winship, Christopher (2010). Effect Heterogeneity and Bias in Main-Effects-Only Regression Models. I: DECHTER, R., GEFFNER, H. & HALPERN, J. Y. (eds.) *Heuristics, Probability and Causality. A Tribute to Judea Pearl* London: College Publications.
- Firebaugh, Glenn (2008). *Seven rules for social research*, Princeton, N.J.: Princeton University Press.
- Frisell, Thomas, Öberg, Sara, Kuja-Halkola, Ralf & Sjölander, Arvid (2012). Sibling Comparison Designs: Bias From Non-Shared Confounders and Measurement Error. *Epidemiology* 23, 713–720
10.1097/EDE.0b013e31825fa230.
- Frölich, Markus (2008). Parametric and Nonparametric Regression in the Presence of Endogenous Control Variables. *International Statistical Review* 76, 214–227.
- Imbens, Guido W. & Angrist, Joshua D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica* 62, 467–475.
- Mohn, Ferdinand A. & Hermansen, Are S. (2013). Kan vi forbedre våre årsakslutninger ved å tegne grafer? En innføring i den kontrafaktiske modellen og kausale grafer (DAGs). *Sosiologi i dag*.
- Morgan, Stephen L. & Winship, Christopher (2007). *Counterfactuals and causal inference: methods and principles for social research*, New York: Cambridge University Press.
- Murray, Michael P. (2006). Avoiding Invalid Instruments and Coping with Weak Instruments. *The Journal of Economic Perspectives* 20, 111–132.
- Murray, Michael P. (2010). The Bad, the Weak, and the Ugly: Avoiding the Pitfalls of Instrumental Variables Estimation. *Working paper*.

- Pearl, Judea (2009). *Causality: models, reasoning, and inference*, Cambridge: Cambridge University Press.
- Pearl, Judea (2010). On Measurement Bias in Causal Inference. I: GRUNWALD, P. & SPIRITES, P. (eds.) *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*. Corvallis, OR: AUAI.
- Skog, Ole-Jørgen (2004). *Å forklare sosiale fenomener: en regresjonsbasert tilnærming*, Oslo: Gyldendal akademisk.
- Vanderweele, Tyler J. & Hernán, Miguel A. (2012). Results on Differential and Dependent Measurement Error of the Exposure and the Outcome Using Signed Directed Acyclic Graphs. *American Journal of Epidemiology*.
- Wooldridge, Jeffrey M. (2006). *Introductory econometrics: a modern approach*, Mason, Ohio: Thomson South-Western.
- Wooldridge, Jeffrey M. (2010). *Econometric analysis of cross section and panel data*, Cambridge, Mass.: MIT Press.

Abstract

Covariate adjustment is the most common approach used to identify causal effects in the social sciences. In cases where one or several confounding variables are unobserved, one or several confounding variables are measured with error, the explanatory variable is measured with error and/or there is reverse causation, using covariate adjustment will result in biased estimates. In this article I present and discuss an alternative approach used to identify causal effects: instrumental variables. Instrumental variables could be used to avoid bias due to correlation between the explanatory variable and the error term, and is therefore a useful tool to identify causal effects. However, finding instruments that are relevant, non-weak, valid, and only have a positive or negative impact on whether individuals take the treatment (monotonicity) is hard, and the bias in the estimate may be large if any of the assumptions fail. The article stress that in order to successfully use instrumental variables, the researcher must be creative and have a thorough understanding of the area of research.

Keywords: Instrumental variables, invalid instruments, weak instruments, instrument relevance, LATE, monotonicity.