

Jacob Gerner Hariri

# Statskundskabens sammenfiltrede virkelighed og et bud på en løsning: IV-estimation

IV-estimation er efterhånden blevet en fast og nødvendig del af den empiriske værktøjskasse i statskundskaben. Den er nødvendig, fordi vores empiriske genstandsfelt er sammenfiltret; variable hænger sammen på kryds og tværs, og det er langtfra ligetil at afgøre, hvad der er årsag til hvad. Formålet med IV-estimation er netop at skære igennem sammenfiltreringen og at afklare, hvorvidt og hvor meget én faktor er årsag til en anden. Denne artikel motiverer og forklarer brugen af instrumentvariable i statistiske analyser, ligesom der gives praktiske råd til analyser med instrumentvariable.

I statskundskaben må de fleste empiriske analyser forlade sig på virkeligheden, som den udspiller sig uden for vinduet. Det er simpelthen ikke alle politologisk relevante problemstillinger, der kan undersøges eksperimentelt – hverken i laboratoriet eller i felten. Derfor studerer vi virkeligheden, og virkeligheden er kompleks: Variable hænger sammen på kryds og tværs, og derfor kan det i mange tilfælde være meget svært at afgøre, hvad der er årsag, og hvad der er virkning.

Det er denne sammenfiltrede virkelighed, hvor  $x$  påvirker  $y$ ,  $y$  påvirker  $x$ , og alt muligt andet påvirker både  $x$  og  $y$ , der motiverer brugen af instrumentvariable. Idéen er simpel: Find en faktor, som kun påvirker  $x$  og derudover ingen forbindelse har til de øvrige variable i den empiriske analyse. Denne faktor kaldes en instrumentvariabel eller simpelthen et instrument. Fordi dette instrument er uafhængigt af  $y$  og alt andet i modellen, kan det isolere variationen i  $x$ , som ikke kommer fra  $y$  (instrumentet er jo ikke forårsaget af  $y$ ), og som heller ikke kommer fra de andre faktorer i modellen. Denne uafhængige eller eksogene variation i  $x$  kan derefter bruges til at identificere den kausale effekt af  $x$  på  $y$ .

Selvom idéen nok er simpel, kan IV-estimation være vanskeligt i praksis. Det er simpelthen svært at finde brugbare instrumenter, faktorer som udelukkende påvirker  $x$  og intet andet i en statistisk model. Af samme grund benytter vi i næste afsnit to illustrative eksempler til at motivere og forklare brugen af IV-estimation. Det ene undersøger effekten af politi på kriminalitet, hvor valgår bruges som instrument for antallet af politifolk i en by (Levitt, 1997).

Det andet undersøger effekten af økonomisk vækst på sandsynligheden for militær konflikt i Afrika syd for Sahara. Her bruges nedbør som instrument for økonomisk vækst. Disse eksempler gør forhåbentlig statistikken håndgribelig. Artiklen fortsætter med en grafisk og siden lidt mere teknisk fremstilling. Afslutningsvis gives et antal praktiske råd til IV-estimation.

## To eksempler på IV-estimation

### *Eksempel 1: politi og kriminalitet*

Steven Levitt forsøgte i 1997 at besvare et tilsyneladende simpelt kausalt spørgsmål: Falder kriminaliteten i en by, hvis der ansættes flere politifolk? Og hvor meget i så fald? Dette er det relevante spørgsmål for politikere, der skal prioritere offentlige midler: Hvor meget kan kriminaliteten forventes at falde for en given bevilling til politiet?

Selvom spørgsmålet er simpelt, er svaret det ikke. På den ene side gælder det, at flere politifolk givetvis begrænser kriminaliteten. På den anden side ansporer kriminalitet antageligt et byråd til at prioritere politiindsatsen; højere kriminalitet giver således mere politi.

Forventningen er derfor i udgangspunktet, at flere politifolk på den ene side er årsag til mindre kriminalitet, mens mere kriminalitet på den anden side er årsag til flere politifolk. En velkendt OLS-regression af kriminalitet på antallet af politifolk indeholder begge sammenhænge, og i Levitts data (som i mange studier før hans) var der faktisk en positiv sammenhæng mellem kriminalitet og antallet af politifolk; selv når der blev kontrolleret for socioøkonomiske og demografiske faktorer, størrelsen af sociale ydelser, uddannelsesniveau arbejdsløshed og andre relevante faktorer. Den positive effekt af kriminalitet på politistyrken dominerer altså den negative effekt af politistyrken på kriminalitet.

I dette tilfælde er det oplagt, at den observerede korrelation ikke kan udlægges som den kausale effekt af politi på kriminalitet; det lyder usandsynligt, at flere politifolk i sig selv skulle være årsag til mere kriminalitet. Lige så oplagt er det, at korrelationen er ubrugelig som baggrund for policy-anbefalinger: Den logiske konsekvens ville være at anbefale at fyre alle politifolk for derved at begrænse kriminaliteten mest muligt.

Steven Levitts IV-analyse tager afsæt i den betragtning, at antallet af politifolk i amerikanske byer ganske ofte stiger i de år, hvor der er valg til byråd og borgmesterpost: Når valget nærmer sig, sættes der flere penge af til politiet. I Levitts data stiger antallet af politifolk da også i gennemsnit med 2 pct. i valgår. Levitt benytter disse elektorale cykler til at isolere den del af variationen i antallet af politifolk, som er uafhængig af graden af kriminalitet i byerne. Da valgperioderne er faste, bliver tidspunktet for valg til byrådene ikke påvirket

af kriminaliteten, og derfor bliver den del af variationen i antallet af politifolk, som kan henføres til valgcyklerne, heller ikke påvirket af kriminaliteten. Levitt bruger altså valgcyklerne som instrumentvariabel for antallet af politifolk, hvilket umiddelbart ser ud til at fjerne problemet med omvendt kausalitet, nemlig at graden af kriminalitet også forårsager variation i politistyrken. Analysen er dog ikke skudsikker. For det kan meget vel tænkes, at der også er valgcykler i fx socialpolitikken eller på uddannelsesområdet. Hvis det er tilfældet (således at det kommunale forbrug på disse områder også vokser i valgår), og hvis disse områder også varierer systematisk med kriminaliteten, ja så vil Levitts analyse overvurdere den kausale effekt af politi på kriminalitet: Den estimerede effekt vil så indeholde effekten af en forbedret socialpolitisk indsats og øgede udgifter til uddannelse, som må formodes også at virke dæmpende på kriminaliteten. Levitt kontrollerer imidlertid for disse faktorer og mener dermed at have isoleret den rene kausale effekt af politi på kriminalitet. For fuldstændighedens skyld runder vi lige eksemplet af og nævner, at ifølge Levitts IV-analyse har antallet af politifolk en betydelig dæmpende effekt på voldskriminalitet men ingen effekt på berigelseskriminalitet.

### *Eksempel 2: nedbør, vækst og militær konflikt*

Antag, som det andet eksempel, at vi ønsker at identificere den kausale påvirkning af økonomisk vækst på sandsynligheden for borgerkrig på tværs af lande i Afrika. Dette er en vanskelig empirisk problemstilling. For det første er der endnu engang omvendt kausalitet, idet sandsynligheden for borgerkrig påvirker væksten, lige så vel som væksten påvirker sandsynligheden for borgerkrig. Dette problem kan ikke løses ved simpelthen at betragte værdier for national indkomst på et tidligere tidspunkt end borgerkrigen: Den nationale indkomst bliver skabt af borgere, som træffer deres beslutning om økonomisk aktivitet på baggrund af deres forventninger om fremtiden – herunder naturligvis deres forventning om sandsynligheden for borgerkrig. For det andet er der næsten med sikkerhed historiske og kulturelle faktorer, som ikke kan observeres, men som hænger sammen med både væksten og sandsynligheden for borgerkrig. Alt dette betyder, at den nationale indkomst ikke er eksogen i denne sammenhæng, og størrelsen af den kausale påvirkning herfra på sandsynligheden for borgerkrig kan ikke umiddelbart identificeres.

Edward Miguel, Shanker Satyanath og Ernest Sergenti (2001) løste problemet ved i et kendt studie at benytte variation i nedbøren som kilde til eksogen variation i den nationale indkomst i afrikanske lande. Da landbrug stadig spiller en betydelig økonomisk rolle i Afrika syd for Sahara, har mængden af nedbør målbare nationaløkonomiske konsekvenser. Derfor benyttede forfat-

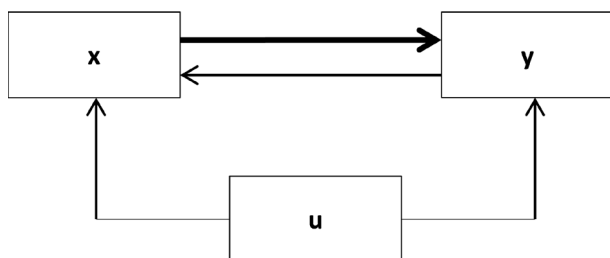
terne kun den del af variationen i væksten, som afhænger af nedbøren, til at identificere den kausale påvirkning af indkomst på sandsynligheden for borgerkrig. Dette giver kausal identifikation, fordi nedbør i denne sammenhæng er eksogen: Sandsynligheden for borgerkrig påvirker ikke mængden af nedbøren, ligesom det nok er begrænset i hvilken grad, at de faktorer som påvirker nedbøren (fx atmosfæriske trykforhold) påvirker sandsynligheden for borgerkrig andet end gennem den økonomiske vækst. Med nedbør som instrument for vækst finder Miguel et al. en meget stærk effekt af økonomisk recession på sandsynligheden for militær konflikt i Afrika syd for Sahara.

I begge disse eksempler er der betydelig omvendt kausalitet, idet den afhængige variabel også påvirker den uafhængige. Desuden er der et væld af faktorer, som påvirker både den uafhængige og den afhængige. Derfor er der intet som *a priori* tilsiger, at politi skulle påvirke kriminalitet, eller økonomisk recession skulle påvirke konflikt snarere end omvendt. Netop derfor er det en væsentlig opgave for empirisk samfundsforskning at forsøge at udrede trådene og identificere så præcist som muligt, hvor meget man kan forvente, at kriminaliteten falder, hvis politistyrken øges, og hvordan økonomien påvirker sandsynligheden for konflikt. Redskabet til at isolere envejspåvirkningen var i begge tilfælde IV-estimation, og i de følgende afsnit udfoldes tankegangen bag metoden. Først grafisk og siden sættes metoden på formel.

## IV-estimation grafisk set

Figur 1 illustrerer politologiens sammenfildrede virkelighed. Der er en sammenhæng mellem to faktorer,  $x$  og  $y$ , men vi kan ikke i udgangspunktet afgøre, hvorfra sammenhængen kommer.

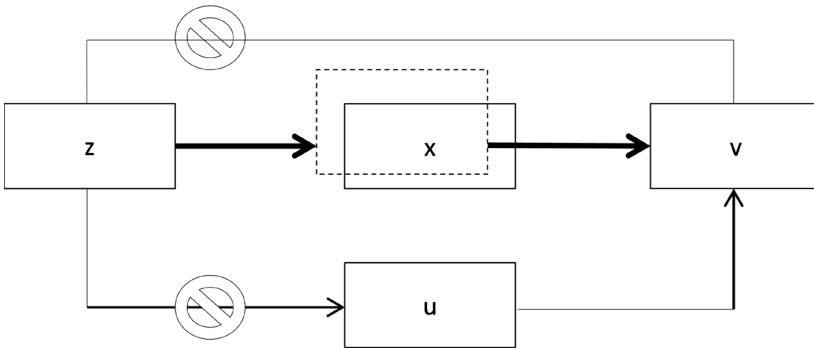
Figur 1: Korrelationsanalyse hvor effekten af  $x$  på  $y$  ikke kan identificeres



Er  $x$  årsag til  $y$ ,  $y$  årsag til  $x$ , forårsages de begge af  $u$ , eller er det i virkeligheden lidt af alle disse sammenhænge på en gang? Hvis vi gerne vil identificere hvor meget,  $x$  påvirker  $y$  – den fede pil i figuren – er det altså ikke nok at betragte

korrelationen mellem de to faktorer. Denne afspejler jo potentielt samtlige pile i figuren. Pointen med IV-estimation er netop at eliminere de forstyrrende pile i figuren, således at forskningsspørgsmålet kan besvares, og effekten af  $x$  på  $y$  kan udsøndres.

Figur 2: IV-estimation



Modsat figur 1 er der i figur 2 to pile. Det illustrerer, at sammenhængen mellem  $x$  og  $y$  ved IV-estimation estimeres i to trin. Det andet trin besvarer det egentlige forskningsspørgsmål (hvor stor er effekten af  $x$  på  $y$ ), mens det første trin udelukkende tjener til at eliminere de pile, der forstyrrede den egentlige analyse. Hvordan så det?

I første trin af analysen estimeres sammenhængen mellem instrumentet og den uafhængige variabel. Instrumentet er illustreret ved kassen  $z$ . Den stiplede firkant, der omfatter en del af  $x$ , illustrerer, at instrumentet aldrig kan opfange al variationen i den uafhængige variabel. Det er dog væsentligt, at instrumentet opfanger en væsentlig del af variationen i den uafhængige, grafisk svarende til at den stiplede kasse så vidt muligt er sammenfaldende med den uafhængige variabel. Hvis instrumentet kun opfanger en begrænset del af variationen i den uafhængige variabel, bliver effekten af den uafhængige på den afhængige variabel estimeret upræcist i analysens andet trin. I det tilfælde siger vi, at instrumentet er svagt. Eller med andre ord: Jo stærkere korrelationen er mellem instrumentet og den endogene uafhængige variabel, desto stærkere instrument og desto mere præcist bliver estimeret i analysens andet (og egentlige) trin.

For at IV-estimation kan løse problemerne med omvendt kausalitet (pilen fra  $y$  til  $x$ ) og udeladte variable (korrelationen mellem  $u$  samt  $x$  og  $y$ ), skal instrumentet opfylde følgende: Det skal ikke have nogen sammenhæng med  $y$  andet end via det  $x$ , som det instrumenterer for. Dette illustreres af de "forbudte" pile

fra  $z$  til  $y$  og fra  $z$  til  $y$  via  $u$ . Hvis instrumentet opfylder dette, er det eksogent (uafhængigt af de andre faktorer i modellen), og analysens første skridt kan således bruges til at isolere den eksogene variation i  $x$ . Hvis instrumentet udelukkende er korreleret med  $y$  via det endogene  $x$ , er instrumentet validt.

Opsummerende kræves det altså, at instrumentet skal være (i) stærkt (stærkt korreleret med det endogene  $x$ , som det er instrument for; illustreret ved sammenfaldet mellem den stiplede kasse og  $x$ ) og (ii) validt (ikke korrelerer med  $y$  andet end via det endogene  $x$ ; illustreret ved de forbudte pile). Hvis disse betingelser er opfyldt, giver IV-estimation et konsistent estimat af den kausale effekt af  $x$  på  $y$ . Det ser vi nærmere på i næste afsnit.

Før vi kommer dertil, skal det nævnes, at særligt den anden betingelse i praktisk arbejde går under forskellige betegnelser. Hvis den anden betingelse gælder, siges det fx også, at instrumentet er ortogonalt på fejleddet, eller at eksklusionsrestriktionen er opfyldt. "Ortogonalt på fejleddet" betyder blot ukorreleret med alt det, vi burde have men ikke har med i modellen. I den anden formulering ligger, at hvis instrumentet udelukkende påvirker  $y$  gennem  $x$ , kan instrumentet udelades fra modellen, når  $x$  er indeholdt. Givet  $x$  har instrumentet altså ingen selvstændig forklaringskraft på  $y$ .

## IV-estimation sat på formel

I sin simpleste form foregår IV-estimation simpelthen som almindelig OLS i to trin og kaldes af samme grund *two-stage least squares* (ofte forkortet 2SLS eller TSLS). I første trin benyttes instrumentet som nævnt til at isolere eksogen variation i den uafhængige variabel, og i andet trin benyttes denne eksogene variation så til at estimere den kausale effekt af den uafhængige variabel på den afhængige variabel.

De to trin i en 2SLS-analyse kan matematisk beskrives som følger:

$$x_i = c + az_i + dX_i + e_i \tag{1}$$

$$y_i = k + bx_i + fX_i + u_i \tag{2}$$

Ligning (2) beskriver analysens *second stage*, som er det egentlige forskningsspørgsmål; nemlig effekten af  $x$  på  $y$ . Denne er givet ved parameteren  $b$ .  $b$  svarer grafisk til pilen længst mod højre i figur 2. OLS-estimation af ligning (2) ville give et skævt resultat (*biased* koefficienter), fordi  $u_i$  og  $x_i$  per antagelse er korreleret. Ligning (1) repræsenterer analysens *first stage* og tjener til at identificere den eksogene variation i  $x$  ved hjælp af instrumentet  $z$ . Parameteren  $a$  angiver korrelationen mellem instrumentet og den endogene uafhængige variabel og svarer grafisk til pilen længst mod venstre i figur 2. Grafisk tjener analysens

første trin altså til at eliminere alle de forstyrrende pile, der indgik i figur 1.  $X$ ' angiver forskellige (eksogene) kontrolvariable, som både indgår i analysens første og andet trin.

IV-estimation giver et konsistent estimat af den kausale effekt af  $x$  på  $y$ , hvis to betingelser er opfyldt. Matematisk ser de ud som følger:

$$\text{Cov}(z, x) \neq 0 \quad (\text{i})$$

$$\text{Cov}(z, u) = 0. \quad (\text{ii})$$

(i) siger, at der skal være sammenhæng mellem instrumentet,  $z$ , og den endogene variabel,  $x$ . Hvis (i) ikke er opfyldt, er instrumentet svagt. (ii) siger, at instrumentet ikke må være korreleret med fejleddet i analysens andet trin, hvilket er det samme som, at instrumentet ikke må være korreleret med den afhængige variabel,  $y$ , andet end gennem  $x$ . Hvis (ii) er opfyldt, er instrumentet validt. Samlet set giver parameteren  $b$  et konsistent estimat af den kausale effekt af  $x$  på  $y$ , når både (i) og (ii) er opfyldt.

Den velkendte OLS-estimator er *unbiased*. Det betyder, at uanset stikprøvestørrelsen vil koefficienterne fra en OLS-analyse i gennemsnit ramme rigtigt i forhold til koefficienterne i befolkningen. Det samme gælder desværre ikke for 2SLS-estimatoren, som ikke er unbiased men dog *konsistent*. Det betyder, at koefficienterne fra 2SLS-analyser nærmer sig befolkningskoefficienterne, jo større stikprøven bliver. 2SLS-estimatoren er ikke unbiased, fordi analysens første trin er resultat af en estimation (i ligning (1)) og dermed indeholder lidt tilfældig variation. En del af denne tilfældige variation kommer fra det endogene  $x$ , som jo er den afhængige variabel i første trin. Og da  $x$  per definition er korreleret med fejleddet i analysens andet trin (ellers ville vi jo ikke behøve at instrumentere), siver en smule af den bias, som oprindeligt motiverede brugen af 2SLS frem for OLS, alligevel ind i estimatoren (Angrist og Pischke, 2009: 209).

I de følgende to afsnit zoomer vi ind på de to betingelser, (i) og (ii). Hvordan ved vi, om de er opfyldt, og hvilke konsekvenser har det for analysen, hvis de ikke er?

## Svage instrumenter?

I figur 2 illustrerede den stiplede kasse, at instrumentet sjældent kan forklare al variationen i den endogene uafhængige variabel. Hvis instrumentet i første trin kun forklarer en mindre del af variationen i  $x$ , bruges kun en lille del af variationen i  $x$  til i analysens andet trin at forklare  $y$ . Og dermed bliver estimatet af effekten af  $x$  på  $y$  upræcist. Instrumentvariable, som kun forklarer en

begrænset del af variationen i det endogene  $x$ , kaldes *svage instrumenter*. Svage instrumenter har to konsekvenser.

For det første forværres den indbyggede bias af 2SLS-estimatoren. 2SLS-estimatoren er som nævnt ikke unbiased, fordi første trin indeholder lidt tilfældig variation i  $x$ , hvoraf noget uvægerligt kommer fra fejlleddet  $u$ . Hvis instrumentet kun forklarer en lille del af  $x$ , kommer en relativt større del af den tilfældige variation i første trin fra  $x$  selv – og dermed fra  $u$ , som  $x$  jo per definition er korreleret med. Derfor bliver 2SLS-estimatoren mere biased, jo svagere instrumentet er. I yderste instans, hvor instrumentet ikke blot er svagt men decideret irrelevant (svarende til  $\text{Cov}(z, x) = 0$ ) og slet ikke bidrager til at forklare  $x$ , svarer 2SLS-estimatoren til OLS-estimatoren. Første trin bortfalder så at sige, så analysen udelukkende består af ligning (2), hvilket i sig selv er en almindelig OLS-regression. Generelt gælder det således, at 2SLS-estimatoren bliver biased i retning af OLS-estimatoren, jo svagere instrumentet er.

For det andet og mere teknisk betyder svage instrumenter, at 2SLS-estimatoren ikke længere nødvendigvis er normalfordelt. Ligesom ved OLS, kræver statistisk inferens i 2SLS-analyser, at estimatoren er normalfordelt i store stikprøver. Derfor kan vi ikke stole på konfidensintervaller og  $p$ -værdier i 2SLS-analyser, når instrumenterne er svage (se fx Stock og Watson, 2007: 440).

Hvornår er et instrument svagt? Tommelfingerreglen siger, at hvis  $F$ -teststørrelsen i analysens første trin er mindre end 10, er instrumentet svagt (Stock, Wright og Yogo, 2002).<sup>1</sup> Der er selvsagt ikke tale om et absolut kriterium, men det anbefales i så fald at undersøge instrumentets robusthed.

Generelt gælder det, at  $F$ -teststørrelsen bliver mindre, hvis man medtager forklarende variable, der er svagt korreleret med den afhængige variabel. Dette gælder også i IV-analysens første trin: Hvis vi lidt forsimpelende antager, at der er to instrumenter og disse er ukorrelerede, så er  $F = \frac{1}{2}(t_1 + t_2)$ ; gennemsnittet af de kvadrede  $t$ -teststørrelser (fx Stock og Watson, 2007: 228). Skrevet på denne form er det klart, at  $F$ -teststørrelsen – og dermed styrken af instrumenterne i første trin – falder, hvis man tilføjer et instrument, som er svagt korreleret med det endogene  $x$ . Implikationen heraf er, at man ikke skal medtage relativt svage instrumenter, hvis man allerede har ét forholdsvist stærkt instrument. Hellere ét stærkt instrument end mange svage!

Hertil kommer at bias i 2SLS alt andet lige er voksende i antallet af instrumenter (Angrist og Pischke, 2009: 209). Faktisk gælder det, at når der præcis er ét instrument per endogen uafhængig variabel, er 2SLS-estimatoren tilnærmelsesvist unbiased. Derfor kan det i tilfælde med relativt svage instrumenter være en god idé at gentage analysen med kun ét instrument per endogen  $x$  (her



skal man naturligvis i hvert tilfælde vælge det stærkeste). Da estimatoren er her tilnærmelsesvist unbiased, minimerer dette problemet med svage instrumenter.

## Holder eksklusionsrestriktionen?

Formålet med IV-estimation er jo i udgangspunktet at isolere den del af variationen i  $x$ , som ikke er korreleret med  $u$ . Men hvis instrumentet selv er korreleret med  $u$ , er det klart, at IV-analysen ikke kan give et konsistent estimat.

Hvor det første krav til IV-estimation var let at undersøge, er det straks værre med det andet. Det andet krav involverer nemlig noget potentielt uobserverbart – fejleddet i analysens andet trin – som ikke må være korreleret med instrumentet. Derfor kan det ikke testes direkte, og her er der i praksis ingen vej uden om simpelthen at tænke sig om. Virker det plausibelt, at instrumentet udelukkende påvirker den afhængige variabel gennem  $x$ ?

Hvis IV-modellen er overdetermineret (flere instrumenter end endogene uafhængige variable), kan man undersøge om ét af instrumenterne er eksogent under antagelse af, at de andre er. Tanken bag den sådanne ”overidentifikations-tests” er at sammenligne IV-koefficienterne fra ét sæt instrumenter med de tilsvarende koefficienter baseret på et andet sæt af instrumenter. Hvis instrumenterne er eksogene, vil disse koefficienter ikke adskille sig meget fra hinanden. Problemet med sådanne tests er, at  $z_1$ 's eksogenitet kun kan testes under forudsætning af, at  $z_2$  faktisk er eksogen. Hvilket vi sjældent kan vide med sikkerhed. Hertil kommer, at hvis nulhypotesen (at instrumenterne er valide) afvises, kan man ikke umiddelbart afgøre, om det er  $z_1$ ,  $z_2$  eller dem begge, som ikke opfylder eksklusionsrestriktionen.

Alternativt kan man medtage sit instrument direkte i analysens andet trin. Hvis instrumentet udelukkende påvirker  $y$  via  $x$ , skal koefficienten på  $z$  være insignifikant, når det optræder sammen med  $x$ . Problemet her er imidlertid, at siden vi i udgangspunktet laver IV, må  $x$  være endogen, hvilket betyder, at alle koefficienter i analysen bliver estimeret med bias (Sovey og Green, 2011: note 4). Derfor kræves det også i dette tilfælde, at modellen er overdetermineret: På grund af endogeniteten skal  $x$  instrumenteres med  $z_2$ , hvis  $z_1$  inkluderes direkte i modellens andet trin.

Der kræver altså altid mere end ét instrument per endogen  $x$ , hvis det skal testes, om eksklusionsrestriktionen holder – og selv hvis det er opfyldt, bør sådanne tests aldrig stå alene. Et overidentifikationsstest kan dog være et udmærket supplement til et godt argument for, hvorfor eksklusionsrestriktionen er overholdt. Det primære bør imidlertid være det gode argument.

## Praktiske råd i IV-analyser

### *Rapporter altid F-teststørrelsen i analysens første trin*

Som nævnt fungerer IV-estimation kun, hvis instrumentet er (tilstrækkeligt) korreleret med den endogene uafhængige variabel. Da dette i praksis oftest måles ved hjælp af F-teststørrelsen i analysens første trin, bør man altid rapportere denne i sin IV-analyse. Til gengæld er  $R^2$ , som måler modellens forklaringskraft, ikke informativ i analysens andet trin.<sup>2</sup> Formålet med IV-estimation er altid at opnå det mest præcise estimat af den kausale effekt af  $x$  på  $y$ ; det er altså en *effects of causes*-analyse. Hvis vi var interesserede i, hvordan en statistisk model som helhed forklarer en afhængig variabel (en *causes of effects*-analyse), er IV-estimation ikke anvendelig, da  $R^2$  ikke har samme naturlige fortolkning som i OLS (fx Wooldridge, 2003: 494).

### *Medtag de samme kontrolvariable i første og andet trin*

De samme variable skal indgå i analysens første og andet trin; 2SLS-estimatoren bliver således inkonsistent, hvis man i analysens første trin kun medtager nogle af de (eksogene) kontrolvariable fra analysens andet trin (fx Wooldridge, 2002: 93). I praksis er dette dog ikke det store problem, da de fleste statistikprogrammer (herunder STATA) af sig selv medtager alle eksogene uafhængige variable i begge analysens trin.

### *Brug ikke endogene kontrolvariable i IV-analyser*

Af det foregående råd følger umiddelbart, at man ikke kan bruge endogene kontrolvariable i 2SLS-analyser. Hvis de medtages i første analyses første trin, er eksklusionsrestriktionen jo ikke opfyldt. Og hvis de ikke gør, er IV-estimatoren som nævnt ikke konsistent. Derfor benyttes der i IV-analyser oftes ret generiske kontrolvariable (såsom regions- og tidsdummier) samt kontroller, hvis eksogenitet er nogenlunde plausibel. I mange sammenhænge kan det alligevel være relevant at kontrollere for endogene faktorer, simpelthen for at få afgjort om den postulerede sammenhæng mellem  $x$  og  $y$  også findes, når en given faktor holdes konstant. Hvis der ikke findes gode instrumenter for disse endogene faktorer, bør man inkludere dem i en almindelig OLS-analyse, som er at foretrække frem for en misspecificeret 2SLS-analyse (jf. fx Hariri, 2012a: 486).<sup>3</sup>

### *"Intuitiv eksogenitet" er ikke "økonometrisk eksogenitet"*

En faktor er endogen i en given sammenhæng, hvis den bliver bestemt inden for den model, som beskriver pågældende sammenhæng. Hvis ikke er den eksogen. Mange geografiske, topografiske eller klimatiske variable er i de fleste samfundsvidenskabelige analyser eksogene i den intuitive forstand, at de ikke

bliver bestemt af samfundsmæssige faktorer. Deraf følger imidlertid ikke, at de nødvendigvis er eksogene i økonometrisk forstand. Således kræver IV-estimation, at instrumentet er ukorreleret med alt (andet end  $x$ ), som potentielt kunne være korreleret med  $y$ . Selvom forekomsten af malaria i en befolkning fx er intuitivt eksogen, er et lands evne til at imødegå og håndtere malaria en funktion af blandt andet økonomisk udvikling og sundhedsmæssig og bureaukratisk infrastruktur. Derfor kan forekomsten af malaria i en given samfundsvidenskabelig analyse godt være økonometrisk endogen. Pointen er her, at geografiske eller andre ”intuitivt eksogene” faktorer ikke per definition er brugbare instrumenter. Det kræver som sagt, at de i en given sammenhæng er ukorrelerede med fejleddet.

### *Pas på med laggede instrumenter*

Hvis værdien på den uafhængige variabel bestemmes på  $t_0$ , og værdien på den afhængige bestemmes på  $t_1$ , kan værdien på den afhængige ikke direkte være årsag til værdien på den uafhængige. Man kunne derfor være fristet til at bruge  $x_{t_0}$  som instrument for  $x_{t_1}$  i analyser af effekten af  $x_{t_1}$  på  $y_{t_1}$ . Det er dog sjældent en god idé. Hvis der er stiafhængighed i det system, der undersøges, vil den laggede værdi af den endogene uafhængige variabel også være endogen. Det er nogenlunde intuitivt, for hvis  $x_{t_1}$  er korreleret med fejleddet, og  $x_{t_0}$  er korreleret med  $x_{t_1}$  (hvilket jo er stiafhængighed), så vil  $x_{t_0}$  også være korreleret med fejleddet. I situationer med stiafhængighed vil laggede instrumentvariable derfor blot skubbe et uløst endogenitetsproblem tilbage i tid, fra  $t_1$  til  $t_0$  (Hariri, 2012b: 192-194).

### *Feedback og retning på bias*

Det er en gængs misforståelse, at OLS-estimatoren bliver biased i numerisk opadgående retning, hvis feedback-effekten fra  $y$  på  $x$  har samme fortegn som effekten fra  $x$  på  $y$ . At feedback-effekten fra  $y$  til  $x$  således lægges oven i effekten fra  $x$  til  $y$ , og OLS-estimatet fanger dem begge. Antag fx at effekten af demokrati på økonomisk udvikling (BNP) er positiv, og omvendt at effekten af BNP på demokrati også er positiv. OLS-estimatet af demokrati på BNP indeholder naturligvis feedback-effekten (BNP på demokrati) – men selvom de begge er positive, er OLS-estimatet ikke nødvendigvis større end IV-estimatet. Retning på bias afhænger nemlig af den relative størrelse af  $x$  på  $y$  og  $y$  på  $x$ . Hvis effekten af BNP på demokrati er mindre end effekten af demokrati på BNP, så vil OLS-estimatet undervurdere den kausale effekt af demokrati på BNP.<sup>4</sup> OLS-estimatoren skal således snarere ses som gennemsnittet af den kausale effekt af  $x$  på  $y$  og  $y$  på  $x$  end summen af de to. Hvis feedback-effekten af  $y$  på  $x$

er numerisk større (end  $x$  på  $y$ ) og i samme retning bliver OLS-estimatet biased i opadgående retning; hvis effekten af  $y$  på  $x$  er numerisk mindre og i samme retning bliver OLS-estimatet biased nedad.

### *"An instrument does not a theory make"*

Instrumentvariablen er en kilde til eksogen variation i den uafhængige variabel; ikke en dybereliggende forklarende variabel (Rodrik, Subramanian og Trebbi, 2004: 153). Og det er som sådan IV-estimatoren skal fortolkes. I en berømt artikel undersøgte Joshua Angrist (1990) de privatøkonomiske konsekvenser af at have tjent i det amerikanske forsvar. Analysen er vanskelig, fordi der potentielt forekommer selvseleksion, således at personer med forholdsvis begrænsede indtægtsmuligheder i det civile selvselekterer ind i militæret. Angrist forsøgte at løse dette problem ved at bruge lodtrækningsnummeret til session (mere specifikt det såkaldte Vietnam draft lottery) som instrument for tjeneste i hæren. Da lodtrækningen er tilfældig, løser det problemet med selvseleksion.

Pointen er her, at Angrists artikel naturligvis ikke introducerer en teori om lodtrækning til session som årsag til individuel indkomst. Lodtrækningen bruges mere beskedent som identifikationsstrategi, og generelt gælder det som sagt, at instrumentet er et middel til at identificere eksogen variation og ikke en forklarende variabel.

### *Pas på med genbrug*

Det er som nævnt i indledningen ikke nogen helt let sag at finde brugbare instrumenter, som korrelerer (stærkt) med  $x$  men ikke med andet i modellen. Derfor ser man jævnligt instrumenter brugt i én sammenhæng blive genbrugt i en lidt anden sammenhæng. Her i Morck og Yeungs (2011: 50) fine formulering af problemet:

A Tragedy of the Commons has led to an overuse of instrumental variables and a depletion of the actual stock of valid instruments for all econometricians. Each time an instrumental variable is shown to work in one study, that result automatically generates a latent variable problem in every other study that has used or will use the same instrumental variable, or another correlated with it, in a similar context. We see no solution to this. Useful instrumental variables are, we fear, going the way of the Atlantic cod.

Her skal det blot understreges, at hvis et instrumentet, som i et tidligere studie var stærkt korreleret med  $x_1$ , nu skal bruges som instrument for  $x_2$ , så kræver det naturligvis, at  $x_1$  og  $x_2$  er fuldstændig ukorrelerede. Dette vil i mange tilfæl-

de ikke være opfyldt, hvorfor eksklusionsrestriktionen ikke er opfyldt. Derfor skal man passe på med og være på vagt over for genbrug af instrumentvariable.

### *Brug gerne IV-estimation som et supplement til OLS*

Hvis eksklusionsrestriktionen ikke er opfyldt, eller instrumentet er svagt, er IV-estimatoren både biased og upræcis. Derfor kan det i mange sammenhænge være en god idé at betragte IV-analysen som én test blandt flere. Som altid afhænger den konkrete empiriske strategi dog af forskningsspørgsmålet. Nogle gange er OLS-analysen decideret ubrugelig (som fx i Levitts analyse af effekten af politistyrken på kriminalitet); her giver det ikke meget mening at betragte OLS- og IV-estimation som komplementære. I sammenhænge hvor man har mere tiltro til OLS, kan IV-analysen bruges som et tjek blandt flere, der undersøger, hvor troværdig den kausale udlægning af resultaterne er.

### *Kilder til eksogen variation (eller: Hvor er de gode instrumenter?)*

De formelle betingelser for IV-estimation er beskrevet ovenfor – hvor det også blev betonet, at det kan være sin sag at finde et godt instrument. Der er ingen kogebofsfremgangsmåde til at finde et godt instrument. Det kan dog anbefales at konsultere litteraturen, der forklarer den endogene uafhængige variabel. Hvis denne fx er førkolonial statsdannelse, kan man konsultere den antropologiske eller arkæologiske litteratur om statens opståen. Lidt mere generelt og til inspiration kan det nævnes, at gængse kilder til eksogen variation er tilfældigt udtrukne tal (lotterier), ”naturen” (fx vejrforhold, geografi, topografi), historiske faktorer eller lovgivningsmæssige kriterier. Et eksempel på det første kunne være at bruge sessionsnummer som instrument for militærtjeneste i analyser, hvor fx indkomst eller holdninger er den afhængige variabel. Det at arbejde i militæret kan sagtens være korreleret med ikke-observerbare personlighedstræk, som også korrelerer med holdninger eller indkomst; militærtjenesten er altså ikke den eneste forskel på soldater og andre, hvorfor en simpel sammenligning giver skæve estimater af effekten af militærtjeneste. Da sessionsnumre udtrækkes tilfældigt er de en god kilde til eksogen variation (Angrist, 1990). Et velkendt eksempel på vejrforhold som kilde til eksogenitet blev givet i eksemplet fra Miguel et al. ovenfor. Et andet kunne være Feyrer og Sacerdote (2009), der bruger vind og strømforhold i Stillehavet og Atlanterhavet som instrumenter for, hvilke øer der blev koloniseret hvornår og af hvem. Administrative eller lovgivningsmæssige diskontinuiteter (fx kommunegrænser, højdekravet for at blive garder, det at børn født lige før/efter nytår starter i skole et år forskudt osv.) kan under visse omstændigheder bruges som instrument for en uafhængig

variabel af interesse. Dette er et såkaldt *fuzzy regression discontinuity*-design, som behandles andetsteds i dette nummer.

### *LATE: Hvad IV-estimatoren måler*

IV-estimatoren måler præcist den kausale effekt af den uafhængige variabel på den afhængige variabel for det subsample, hvor værdien på den uafhængige variabel bestemmes af instrumentet. Dette subsample er velafgrænset og identificerbart, hvis instrumentet er en binær variabel. I sprogbrugen fra den nyere *treatment*-litteratur kaldes dette en *Local Average Treatment Effect*, forkortet LATE. Antag fx at vi vil estimere afkastet af uddannelse: Hvor meget stiger lønnen i gennemsnit for forskellige uddannelser. Uddannelse og indkomst er korreleret med fx evner og interesser, som er vanskelige at inkludere som kontrolvariable. Et muligt instrument kunne da være, om folk bor i en by med universitet eller ej. IV-estimatoren giver os her effekten af uddannelse på indkomsten, for de mennesker som begynder at studere på universitetet, fordi de bor i en universitetsby. Hvis instrumentet ikke er binært (men i eksemplet fx afstand i kilometer), giver IV-estimatoren os ikke et præcist kausalt estimat for et veldefineret subsample (Blundell og Costa Dias, 2009: 612). Dog gælder det som ovenfor anført, at jo bedre instrumentet i gennemsnit forklarer den uafhængige variabel (jo større F-teststørrelse i analysens første trin), jo mindre lokal og tættere er IV-estimatoren på at gælde hele stikprøven. I grænsen hvor  $R^2$  er 1 i analysens første trin, er instrumentet jo sammenfaldende med den uafhængige variabel, således at IV-estimation, hvis antagelserne holder, giver den gennemsnitlige kausale effekt for hele stikprøven.

## Opsamlende

Estimation ved hjælp af instrumentvariable er efterhånden en uundværlig del af den kvantitative værktøjskasse i empirisk samfundsforskning, hvor det er notorisk vanskeligt at afgøre, hvorvidt og hvor meget en faktor er årsag til en anden. Derfor har artiklen forsøgt at give en ikke-teknisk introduktion til og motivation for brugen af instrumentvariable.

Selvom idéen bag IV-estimation måske er simpel, er det i praksis meget vanskeligt at finde overbevisende instrumentvariable. Lidt karikeret kan man sige, at hvis eksklusionsrestriktionen er opfyldt, er instrumentet formentlig svagt. Og hvis instrumentet er stærkt, er eksklusionsrestriktionen formentlig ikke overholdt. Eller med andre ord vil de faktorer, som udelukkende påvirker  $y$  via  $x$ , ofte være så perifære i forhold til den samfundsfaglige problemstilling, at de heller ikke er specielt stærkt korreleret med  $x$ . Og hvis de er stærkt korreleret med  $x$ , hænger de formentlig også sammen med alt muligt andet i modellen.

Derfor bør estimation med instrumentvariable heller aldrig stå alene i en empirisk undersøgelse – og gør det sjældent i praksis. IV-estimation skal snarere ses som én blandt flere analyser. Hvis instrumentet er svagt, forstærkes nemlig den indbyggede bias ved IV-estimation. Og hvis eksklusionsrestriktionen ikke er opfyldt, er IV-estimatet både skævt og upræcist sammenlignet med mere almindelige metoder. IV-estimation er dog stadig en uhyre brugbar metode i empirisk samfundsforskning, for den tvinger forskeren til at fokusere på den fundamentale betingelse for kausal identifikation; nemlig hvorvidt den uafhængige variabel kan siges at være eksogen. IV-estimation tvinger så at sige samfundsforskeren til at stille de rigtige – og kritiske – spørgsmål til sin egen analyse. Samtidig gør gode IV-analyser os klogere på effekterne af de årsager, vi i politologien interesserer os for. Og præcis identifikation af, hvilke effekter en given faktor (her årsag) har, er en forudsætning for at designe og evaluere policy-tiltag i praksis. Sine svagheder til trods er der derfor ingen tvivl om, at IV-estimation er et blivende redskab i den empiriske politologis værktøjskasse.

## Noter

1. F-teststørrelsen i første trin er nogenlunde omvendt proportional med den relative bias af 2SLS-estimatoren i forhold til OLS-estimatoren. Et F på 10 svarer altså nogenlunde til en relativ bias på 10 pct.
2. Faktisk kan  $R^2$  blive negativ i 2SLS-analyser (se fx Wooldridge, 2003: 503).
3. Formelt set skal instrumentet blot være eksogent betinget af kontrolvariablene i analysen. Der kan dog der være grund til skepsis, hvis eksklusionsrestriktionen kun hævdes at holde, når der er medtaget flere kontrolvariable. Kontrolvariablene skal jo alle selv være økonometrisk eksogene.
4. Her ser vi bort fra udeladte variable, som naturligvis også kan skabe bias i OLS-estimatet. Hvis en udeladt variabel er positivt korreleret med både  $x$  og  $y$ , skaber dette opadgående bias i OLS-estimatet.

## Litteratur

- Angrist, Joshua D. (1990). Lifetime earnings and the Vietnam era draft lottery: Evidence from social security administrative records. *American Economic Review* 80: 313-336.
- Angrist, Joshua D. og Jörn-Steffen Pischke (2009). *Mostly Harmless Econometrics*. Princeton, MA: Princeton University Press.
- Blundell, Richard og Monica Costa Dias (2009). Alternative approaches to evaluation in empirical microeconometrics. *The Journal of Human Resources* 44 (3): 565-640.
- Feyrer, James og Bruce Sacerdote (2009). Colonialism and modern income: Islands as natural experiments. *Review of Economics and Statistics* 91 (2): 245-262.

- Hariri, Jacob Gerner (2012a). The autocratic legacy of early statehood. *American Political Science Review* 106 (3): 471-494.
- Hariri, Jacob Gerner (2012b). Kausal inferens i statskundskaben. *Politica* 44 (2): 184-201.
- Levitt, Steven D. (1997). Using electoral cycles in police hiring to estimate the effect of police on crime. *American Economic Review* 87 (3): 270-290.
- Miguel, Edward, Shanker Satyanath og Ernest Sergenti (2001). Economic shocks and civil conflict: An instrumental variables approach. *Journal of Political Economy* 112 (4): 725-753.
- Morck, Randall og Bernard Yeung (2011). Economics, history, and causation. *Business History Review* 85: 39-63.
- Rodrik, Dani, Arvind Subramanian og Francesco Trebbi (2004). Institutions rule: The primacy of institutions over geography and integration in economic development. *Journal of Economic Growth* 9: 131-165.
- Sovey, Allison og Donald Green (2011). Instrumental variables estimation in political science: A readers' guide. *American Journal of Political Science* 55 (1): 188-200.
- Stock, James H., Jonathan H. Wright og Motohiro Yogo (2005). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics* 20: 518-529.
- Stock, James H. og Mark W. Watson (2007). *Introduction to Econometrics*. New York: Pearson.
- Wooldridge, Jeffrey M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: The MIT Press.
- Wooldridge, Jeffrey M. (2003). *Introductory Econometrics*. South-Western: Mason, OH.