



Kausale design – Et gjennombrudd i samfunnsfagene eller keiserens nye klær?

Causal design – a breakthrough in social sciences, or the emperor's new clothes?

Kristoffer Rypdal

Professor emeritus, Institutt for matematikk og statistikk, UiT – Norges arktiske universitet
kristoffer.rypdal@uit.no

Statistiske metoder som regresjonsanalyse og kausale design basert på naturlige eksperimenter har etter hvert blitt hyllevare innen de fleste fag som behandler store datamengder, ikke minst innen epidemiologi, meteorologi og klimavitenskap.¹ Et eksempel er prinsipal komponentanalyse (PCA), som brukes til å finne karakteristiske mønstre i geofysiske felt. Mønstrene, eller prinsipalkomponentene, tolkes gjerne som uttrykk for karakteristiske dynamiske moder i de globale værssystemene. Disse tolkningene tillegges imidlertid vanligvis liten vekt hvis man ikke samtidig identifiserer de samme modene i dynamiske vær- eller klimamodeller og forstår de viktigste fysiske mekanismene som ligger bak. Observasjonene kan være for dårlige eller sparsomme, eller systemet kan være så komplekst at prinsipalkomponentene ikke representerer noe som gir fysisk mening for oss.

Menneskelige samfunn er langt mer komplekse enn klimasystemet og unndrar seg modellering basert på naturvitenskapelig reduksjonisme. Men både i samfunnsfagene og i studiene av komplekse naturlige systemer er det plass for enkle, konseptuelle dynamiske modeller. Disse er imidlertid bare av verdi hvis man klarer å identifisere de viktigste driverne av endring, og filtrerer bort de uviktige. Kausale design kan være av stor verdi i denne seleksjonsprosessen, men det forutsetter at man ikke bruker dem primært for å «filtrere inn» mekanismer som man av en eller annen grunn tror er viktig, men at man heller bruker dem til å «filtrere ut». I motsatt fall kan bruk av vanskelig tilgjengelige kvantitative metoder lett bli et middel til å stilne protester fra dem som kan ha gode, innsiktsfulle motforestillinger til en gitt hypotese.

1. Jeg er utdannet som fysiker og anvendt matematiker og er selvsagt positivt innstilt til bruk av kvantitative metoder i samfunnsvitenskapene. Jeg har ingen formell skoleing i statistikk og økonometri, og slett ingen bakgrunn i samfunnsfag, men har siden årtusenskiftet jobbet tverrfaglig med klimamodellering og energispørsmål og under pandemien fusket litt med epidemimodeller.

«Brent jord»: et aktuelt eksempel

Utgangspunktet for dette debattinnlegget er en debatt i avisa Nordlys², som oppsto etter publikasjonen av artikkelen til Buck et al. (2022) i *Tidsskrift for samfunnsforskning* med tittelen «Brent jord: langsiktige politiske konsekvenser av nedbrenning og tvangsevakuering i Nord-Troms og Finnmark». Artikkelen, heretter referert til som BMMS, har vært utsatt for kritikk fra historikere, journalister og samfunnsdebattanter. Mye av denne kritikken har dreid seg om historisk kildebruk og i liten grad på bruken av særskilte økonometriske metoder, såkalte kausale design, som jeg oppfatter som et hovedpoeng i arbeidet. Ifølge BMMS anvender de to slike metoder i artikkelen, *endringer-i-endringer* (differences-in-differences, DiD) og den *syntetiske kontrollmetoden* (synthetic control method, SCM), og mener å kunne påvise en kausal forbindelse mellom nedbrenningen av kommuner i Finnmark og Nord-Troms og sviktende valgoppslutning og stemmetall til Arbeiderpartiet (AP) etter 1970.

I realiteten er det ikke SCM-metoden forfatterne anvender, men en versjon av *lineær faktor-modellen*, som anvendes av Xu (2017). Forvirringen oppstår antagelig fordi Xu kaller sin metode *generalized synthetic control* (GSC). Både GSC og SCM konstruerer et kontrafaktisk tidsforløp av valgresultat for en eller flere kommuner som har vært utsatt for nedbrenning, for det hypotetiske tilfellet at de ikke hadde vært brent. For dette formålet brukes en kontrollgruppe av kommuner som ikke har vært utsatt for nedbrenning. For øvrig er metodene svært ulike; SCM konstruerer en estimator for det kontrafaktiske forløpet ved hjelp av et vektet middel av forløp i kontrollgruppen (Abadie, 2021), mens GSC-estimatoren er en lineær kombinasjon av uobserverte faktorer som også konstrueres ved hjelp av kontrollgruppens data, og som gis nye vekter ved å tilpasses til eksperimentgruppens data før intervensjonen (Xu, 2017).

BMMS uttrykker store forhåpninger til å anvende disse metodene i samfunnsvitenskapelig og historisk forskning. En av betenkelighetene er imidlertid at metodene er vanskelig tilgjengelige for mange, også for meg som ikke er økonometriker, og er lett å bruke og fortolke feil. Et svært vanlig fenomen, som eksempelvis er årsaken til så mange motstridende resultater i medisinsk forskning, er *bekreftelsestendensen*, der bekreftelsen bevisst eller ubevisst er innbakt i designet av den statistiske modellen.

Fallgruver ved bruk av DiD-metoden

DiD-metoden er et spesialtilfelle av en multippel, lineær regresjonsmodell som gir en prediksjon av en responsvariabel Y , for eksempel prosentvis valgoppslutning om AP i kommuner identifisert med indeksen k og ved valgåret identifisert ved indeksen t . Tallparet (k, t) identifiserer altså et observert valgresultat $Y_{k,t}$, og regresjonsmodellen gir en prediksjon av dette resultatet ved hjelp av en lineær kombinasjon av fire kjente prediktorer; 1, μ_t , κ_k , $\delta_{k,t}$. I DiD-modellen har prediksjonen følgende form:

$$\hat{Y}_{k,t} = a + b_1\mu_t + b_2\kappa_k + b_3\delta_{k,t}$$

Det spesielle her er at prediktoren μ_t bare avhenger av tiden t , og er den samme for alle kommuner k , og at κ_k bare avhenger av kommunenummer, men er uavhengig av valgår. Foreløpig har vi ikke spesifisert hva prediktorene er, men vi må uansett erkjenne at en slik

2. <https://www.nordnorskdebatt.no/nedbrenningens-lange-skygger/o/5-124-178771>
<https://www.nordnorskdebatt.no/pussig-forklaring-om-brent-jord-og-politiske-konsekvensar/o/5-124-178696>
<https://www.nordnorskdebatt.no/forskningens-ekkokammer/o/5-124-179438>
<https://www.nordnorskdebatt.no/hvorfor-ikke-litt-mer-nysgjerrig-niemi-og-tjelmeland/o/5-124-180183>

enkel modell ikke kan gi en nøyaktig prediksjon. Det knytter seg derfor stor interesse til feilen i prediksjonen $e_{k,t} = Y_{k,t} - \hat{Y}_{k,t}$, som bør være en tilfeldig støy hvis man skal kunne gi en meningsfull fortolkning av modellens resultater.

Formålet med denne typen analyser er å undersøke effekten av en intervensjon (i dette tilfellet nedbrenning) ved å dele populasjonen av kommuner opp i to grupper: én eksperimentgruppe, som er utsatt for intervensjonen, og én kontrollgruppe som ikke er det. Det vanlige er derfor å velge $\kappa_k = 1$ hvis k representerer en nedbrent kommune, og $\kappa_k = 0$ ellers. Koeffisienten b_2 representerer altså en eventuell systematisk forskjell i $Y_{k,t}$ mellom de to gruppene som ikke har sammenheng med intervensjonen. En prediktor som bare har to mulige utfall for en gitt observasjon, kalles en «dummy». Prediktoren $\delta_{k,t}$ er også en dummy som er 0 bestandig, bortsett fra i eksperimentgruppen etter intervensjonen, da den har verdien 1. Koeffisienten b_3 er altså et mål på effekten av intervensjonen.

Leddet $b_1\mu_t$ er ment å modellere den endringen (felles trend) av $Y_{k,t}$ i begge grupper som er uavhengig av intervensjonen. Før og etter intervensjonen er dette leddet det eneste som endrer seg med tiden, og er i modellen den samme for alle kommuner. Modellen gir altså at $\hat{Y}_{k,t}$ utvikler seg parallelt i de to gruppene før og etter intervensjonen, mens det er en parallellforskyvning, en diskontinuitet i $\hat{Y}_{k,t}$ av størrelse b_3 , ved intervensjonstidspunktet. Dette er en grunnleggende antagelse i DiD (paralleltrendhypotesen). Hvis denne antagelsen stemmer dårlig med observasjonsdataene, vil feilen $e_{k,t}$ ikke framstå som en tilfeldig støy, men inneholde trender som ikke er felles for kommunene. Det vil vanskeliggjøre fortolkningen av b_3 som den eneste effekten av intervensjonen. Vi skal også se nedenfor at det vanskeliggjør beregning av usikkerheten i estimatene av b_3 .

BMMS anvender den enkleste versjonen av DiD-modellen, der μ_t velges som en dummy som er 0 før intervensjonen og 1 etter. I denne versjonen predikerer modellen bare én verdi før intervensjonen, og én etter, for hver av de to gruppene. I det følgende kaller jeg den «dummytrendmodellen».

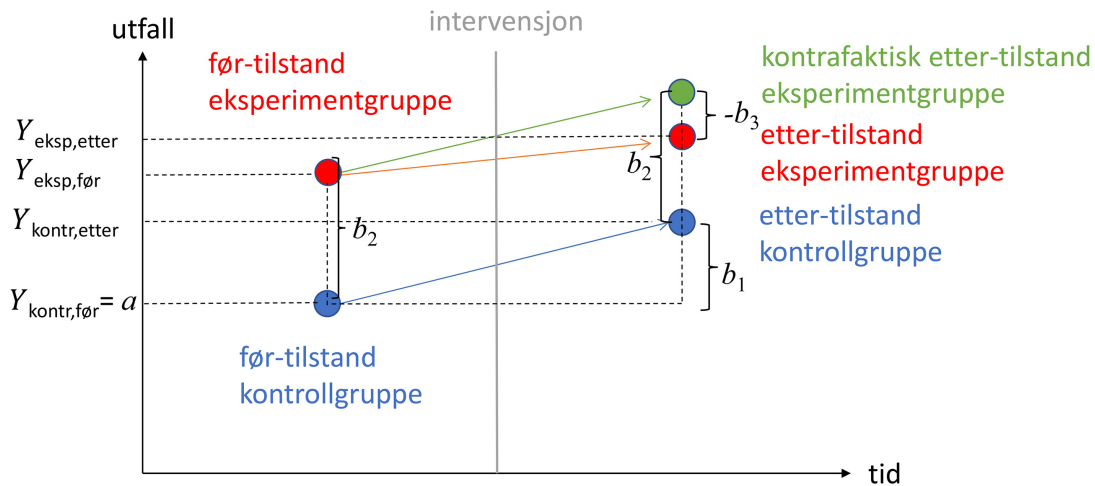
Figur 1 illustrerer hvordan DiD-metoden fungerer i BMMS' versjon. $Y_{\text{eksp,før}}$ er verdien i eksperimentgruppen midlet over alle kommuner i denne gruppen og over alle valg før krigen. $Y_{\text{eksp,etter}}$ er tilsvarende etter krigen. $Y_{\text{kontr,før}}$ og $Y_{\text{kontr,etter}}$ er tilsvarende for kontrollgruppen. Endringen i eksperimentgruppen er (legg merke til at b_3 er negativ i dette eksemplet);

$$\Delta Y_{\text{eksp}} = Y_{\text{eksp,etter}} - Y_{\text{eksp,før}} = (a + b_1 + b_2 + b_3) - (a + b_2) = b_1 + b_3,$$

og endringen i kontrollgruppen er

$$\Delta Y_{\text{kontr}} = Y_{\text{kontr,etter}} - Y_{\text{kontr,før}} = (a + b_1) - (a) = b_1.$$

Hvis forskjellen i endringen mellom eksperimentgruppen og kontrollgruppen (endringer-i-endringer) tolkes som en effekt av intervensjonen, er denne effekten $\Delta Y_{\text{eksp}} - \Delta Y_{\text{kontr}} = b_3$.

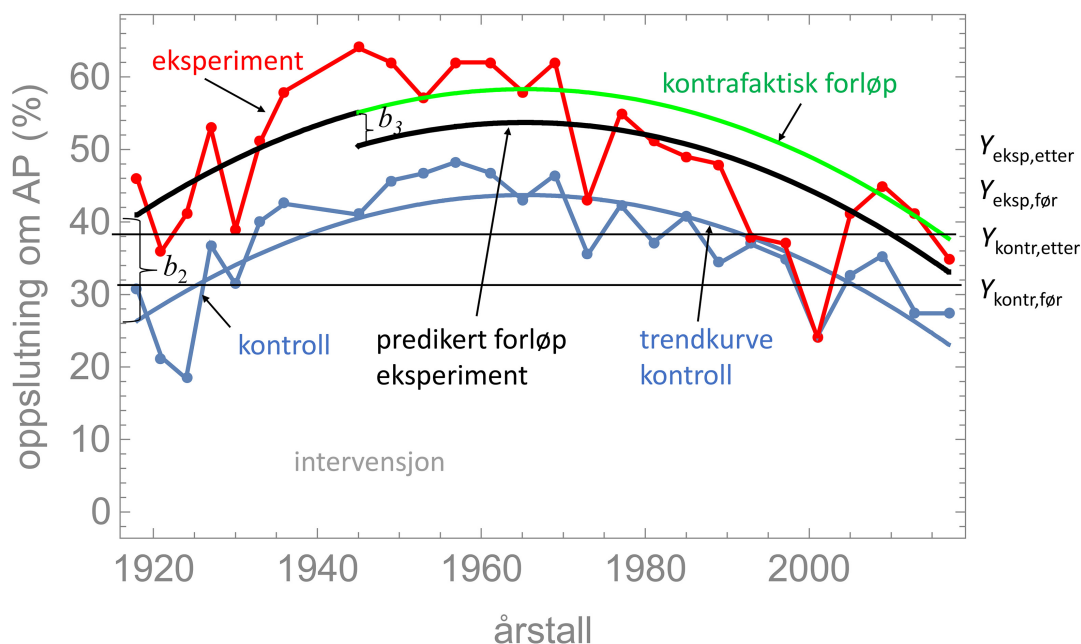


Figur 1. Illustrasjon av prinsippet bak DiD, slik metoden anvendes i BMMS.

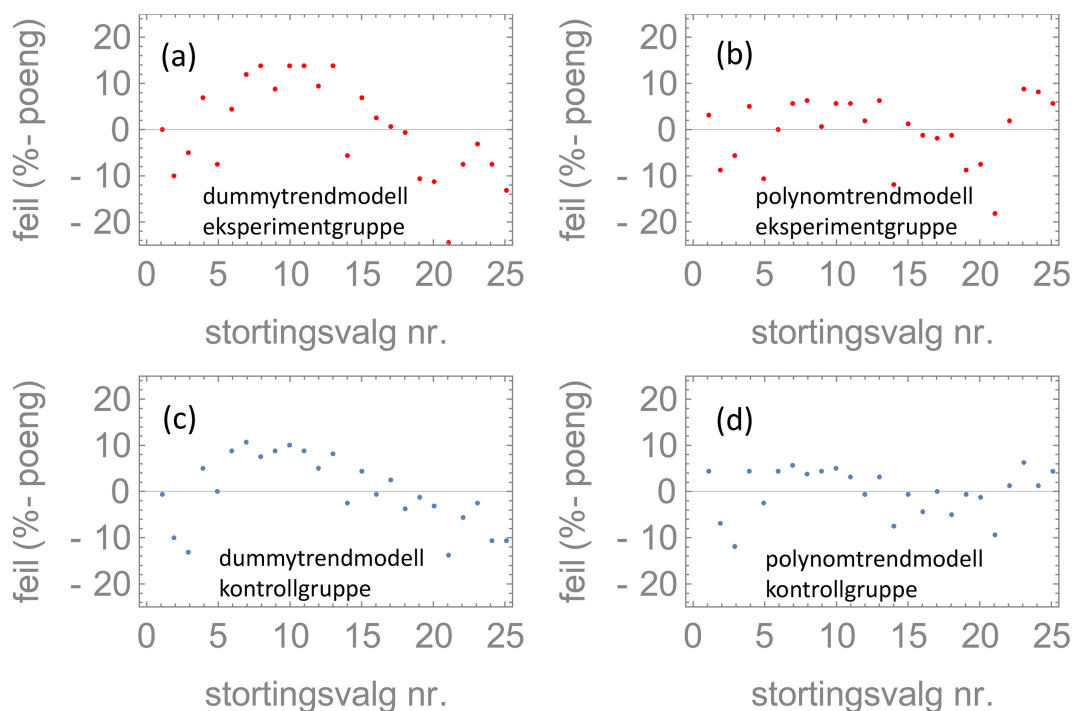
Resultatet av å la μ_t være en dummy i regresjonsmodellen blir at modellen kun predikerer fire verdier for Y : én for hver av de to kommunegruppene før intervensjonen og én for hver gruppe etter intervensjonen (de røde og blå sirklene i figur 1). Disse verdiene er de observerte verdiene midlet over hver gruppe og over valgene henholdsvis før og etter. At modellen er så grov at den predikerer bare fire verdier, gjør at avvikene $e_{k,t}$ mellom predikerte og observerte data blir svært store, og at de inneholder mye informasjon som ikke er tilfeldig støy. Dette framgår av valgresultatene som funksjon av tid, $Y_{\text{eksp},t}$ midlet over eksperimentgruppen og $Y_{\text{kontr},t}$ midlet over kontrollgruppen, som BMMS har plottet i sin figur 2, og som er illustrert i min figur 2 nedenfor. Det er en sterkt voksende trend både i valgdeltagelse og oppslutning om AP i alle grupper før krigen, en flat trend i valgdeltagelse i alle grupper etter krigen, men sterkt fallende trender i oppslutningen om AP som starter to tiår etter krigens slutt.

Figur 2 illustrerer at feilen i prediksjonene kan reduseres ved å erstatte dummyen μ_t med en analytisk, kontinuerlig funksjon (for eksempel et polynom) slik at $a + b_1\mu_t$ blir en tilpasning til tidsrekken for kontrollgruppen. Her er funksjonen et andregrads polynom som er tilpasset ved hjelp av minste kvadraters metode. Jeg kaller denne for «polynomtrendmodellen», som nå vil predikere en tidsrekke med ett utfall $\hat{Y}_{\text{kontr},t}$ for hvert stortingsvalg i kontrollgruppen og en tidsrekke $\hat{Y}_{\text{eksp},t}$ for eksperimentgruppen. Den blå, glatte kurven i figuren er prediksjonen for kontrollgruppen, og den svarte, glatte kurven, med en diskontinuitet ved krigens slutt, er prediksjonen for eksperimentgruppen. Diskontinuiteten i den svarte kurven (b_3) representerer effekten av intervensjonen.

At dette reduserer den midlere kvadratiske feilen i prediksjonene, og de serielle korrelasjonene mellom dem, kan man se fra figur 3. I figur 3(a) har jeg plottet feilene i eksperimentgruppen for dummytrendmodellen, der feilene er avstandene mellom observasjonspunktene og de faste middelverdiene før og etter intervensjonen. I figur 3(b) viser jeg feilene i eksperimentgruppen for polynomtrendmodellen, der disse er avstanden mellom observasjonspunktene og de glatte kurvene i figur 2. Figur 3(c) og (d) viser tilsvarende for kontrollgruppens data. At den nye modellen ikke fjerner de serielle korrelasjonene mellom feilene fullstendig, skyldes hovedsakelig at et andregrads polynom ikke er en perfekt funksjon for tilpasning til kontrollgruppens data, slik at det fortsatt er trender som reflekteres i feilene.



Figur 2. Prinsippskisse av polynomtrendmodellen. De røde punktene representerer oppslutning om AP for de enkelte valgår midlet over kommunene i nedbrente kommuner, og de blå punktene representerer tilsvarende for Sør-Norge (kontrolldata). De horisontale, stiplede linjene angir midlere valgoppslutning om AP i eksperimentgruppen henholdsvis før og etter krigen, og de horisontale heltrukne linjene angir tilsvarende for kontrollgruppen. Den blå, glatte kurven er polynomtilpasningen til kontrolldata, den grønne kurven er det kontrafaktiske forløpet, og den svarte kurven er det modellerte tidsforløpet. Effekten av nedbrenningen antas å være differansen mellom det modellerte og det kontrafaktiske forløpet.



Figur 3. (a) viser feilene $e_{k,t}$ i eksperimentgruppen for alle stortingsvalg fra 1918 til 2017 (unntatt 1945) i dummytrendmodellen som er brukt av BMMS. (b) viser tilsvarende for polynomtrendmodellen. (c) og (d) viser det samme for kontrollgruppen.

Den systematiske tidsavhengigheten av feilledet $e_{k,t}$ (serielle korrelasjoner) har betydning for statistisk signifikans av den den effekten av intervensjonen som detekteres gjennom estimatet av regresjonskoeffisienten b_3 . Det framgår av det foregående at regresjonskoeffisientene a , b_1 , b_2 , b_3 i dummytrendmetoden er bestemt av middelverdiene i de fire gruppene som er angitt som de røde og blå sirklene i figur 1. BMMS estimerer ikke koeffisientene direkte ved å bruke middelverdiene i de fire gruppene, men ved å bruke hele datasettet og finne koeffisientene som minimaliserer kvadratsummen av feilene $\sum_{k,t} e_{k,t}^2$. Fordelen med dette er at det finnes standard metoder som samtidig estimerer usikkerheten i estimatet (for eksempel standardavvik eller 95 % konfidensintervall). Disse usikkerhetene angir forfatterne i figur 3 og 4 og i appendiks 2 sammen med andre statistiske data. I alle analysene de gjør, finner de at verdien null ligger utenfor 95 % konfidensintervall omkring beste estimat av b_3 , og konkluderer derfor med at sannsynligheten for at den sanne verdien av b_3 er null, er mindre enn 5 % ($p < 0.05$). Null-hypotesen om at intervensjonen ikke har noen effekt, er derfor forkastet med p -verdi mindre enn 0.05. I standard statistisk terminologi er dette synonymt med at effekten er statistisk signifikant. Et åpenbart problem er imidlertid at de også finner en signifikant b_3 (altså en signifikant effekt av nedbrenningen) i det tilfellet at eksperimentgruppen velges å være de nordnorske kommunene som ikke ble brent (heretter kalt *nord*), og kontrollgruppen er de sørnorske kommunene (heretter kalt *sør*). Dette framgår av deres figur 3 og 4.

Problemet med denne framgangsmåten er at disse estimatene av standardfeilen i regresjonskoeffisientene er gyldige bare under restriktive antagelser om de statistiske egenskapene til feilvariabelen $e_{k,t}$. Blant de viktigste er at de skal være statistisk uavhengige og identiske (ha samme forventningsverdi og varians). Det er opplagt at dette ikke er tilfellet her; trendene både før og etter intervensjonen gjør at avvikene fra middelverdiene avhenger sterkt av tiden (positive serielle korrelasjoner), og små kommuner har større varians i valgresultater enn store. Det er velkjent at positive serielle korrelasjoner vil gi opphav til lavere estimerte standardfeil for regresjonskoeffisienter enn de sanne feilene. Dette betyr at man ikke kan utelukke at de sanne konfidensintervallene er så vide at de omfatter $b_3 = 0$, og at man dermed ikke kan forkaste null-hypotesen om at intervensjonen ikke har noen effekt. Dette er også hovedkonklusjonen i en systematisk studie av feilkilder i DiD-metoden (Bertrand et al., 2004).

Det framgår av figur 2 og figur 3 at polynomtrendmodellen fjerner en del av de korrelasjonene som skyldes felles trender for eksperiment- og kontrollgruppen, men selvsagt ikke korrelasjonene som skyldes ulikhetene i disse trendene. Dette er en svakhet også fordi disse ulikhetene kan være en effekt av intervensjonen og ikke hører hjemme i feilledet. Problemet oppstår fordi modellen tar ut hele effekten av intervensjonen som en diskontinuerlig endring ved intervensjonstidspunktet (diskontinuiteten i den svarte kurven i figur 2), og ikke tillater en forsinket effekt som kan tilta gradvis over tid. Ett av formålene med GSC er å rette på dette.

GSC-metoden i korte trekk

Modellen som er illustrert i figur 2, er egentlig et spesialtilfelle av lineær faktor-modellen som GSC-metoden bygger på. Det vesentlig nye i GSC er at trendkurven som er antatt til stede både i kontrollgruppen og eksperimentgruppen, blir erstattet av et antall uobserverbare faktorer som avhenger av tiden; $f_t^{(j)}$, $j = 1, \dots, J$. Responsene $\lambda_k^{(j)}$ på disse faktorene, «factor loadings», er ulike for de ulike kommunene, men uavhengige av tiden. Slik blir «trenden» for kommune nr. k gitt ved summen $\sum_{j=1}^J \lambda_k^{(j)} f_t^{(j)}$. Den samme faktoren $f_t^{(j)}$ gjør

seg altså gjeldende i alle kommuner, men med forskjellig styrke $\lambda_k^{(j)}$. Dette betyr at modellen ikke lenger forutsetter parallelle trender i de to gruppene, og det muliggjør at den kraftigere fallende trenden i oppslutningen om AP i eksperimentgruppen kan være et resultat av én eller flere faktorer som beskriver en slik trend, og der responsen $\lambda_k^{(j)}$ er større i eksperimentgruppen enn i kontrollgruppen. Modellen forutsetter altså ikke at ulikhetene i trendene er en konsekvens av intervensjonen. Lineær faktor-modellen tillater også at man legger til en lineær kombinasjon av N ekstra prediktorer (kovariater) i estimatoren; $\sum_{n=1}^N \beta_n x_{k,t}^{(n)}$, på samme måte som i enhver lineær, multippel regresjonsmodell. Dette åpner for muligheten til å veie effekten av intervensjonen mot andre kjente årsaksmekanismer, som kan være ulike for ulike kommuner. Den fulle modellen har formen;

$$\hat{Y}_{k,t} = b_{k,t} \delta_{k,t} + \hat{Y}_{k,t}^{(0)}, \text{ der } \hat{Y}_{k,t}^{(0)} = \sum_{n=1}^N \beta_n x_{k,t}^{(n)} + \sum_{j=1}^J \lambda_k^{(j)} f_t^{(j)},$$

Her er $\delta_{k,t}$ fortsatt 1 for kommuner som har vært utsatt for nedbrenning, og 0 ellers. Parameterne $b_{k,t}$ erstatter b_3 i DiD-modellen og representerer intervensjonseffekten i kommune k ved tiden t . Intervensjonseffekten kan altså både være ulik i ulike kommuner og variere med tiden. $\hat{Y}_{k,t}^{(0)}$ er følgelig prediksjonen uten intervensjon, som for kommuner i eksperimentgruppen og etter intervensjonstidspunktet representerer prediksjonen av det kontrafaktiske forløpet.

Bidragene til prediksjonen fra kovariatene $x_{k,t}^{(n)}$ og faktorene $f_t^{(j)}$ har begge som formål å forklare mest mulig av de ikke-tilfeldige variasjonene i den observerte $Y_{k,t}$, og dermed å redusere feilledet $e_{k,t} = Y_{k,t} - \hat{Y}_{k,t}$ til tilfeldig støy. BMMS innfører ingen kovariate prediktorer i tillegg til nedbrenningen som kan være med på å forklare variasjonene i valgresultatene, og overlater derfor all modellering av de ikke-tilfeldige variasjonene til de uobserverte faktorene. I det følgende skal jeg derfor anta at vi ikke har kovariater i modellen.

Xu (2017) anviser en tretrinns estimeringsprosedyre der først $f_t^{(j)}$ og $\lambda_k^{(j)}$ i kontrollgruppen estimeres fra tidsrekkene i kontrollgruppen for hele perioden både før og etter intervensjonen. Dette gjøres ved å minimalisere midlere kvadratiske avvik mellom observasjoner og prediksjon. De resterende $\lambda_k^{(j)}$, de i eksperimentgruppen, estimeres ved å minimalisere midlere kvadratiske avvik mellom observasjoner og prediksjon i eksperimentgruppen før intervensjonen, gitt estimatene i første trinn. Antallet J av faktorer er gjenstand for subjektivt valg, selv om det finnes ulike prosedyrer for et slikt valg etter ulike kriterier. Det må pålegges føringer på formen til $f_t^{(j)}$ og $\lambda_k^{(j)}$ for at minimaliseringsprosedyren for det kvadratiske avviket skal ha entydig løsning. Her velger man at vektorene $(f_1^{(j)}, f_2^{(j)}, \dots, f_T^{(j)})$ for ulike faktorer j skal være ortonormale, og at $(\lambda_1^{(j)}, \lambda_2^{(j)}, \dots, \lambda_K^{(j)})$, skal være ortogonale. Her er T antall elementer i tidsrekken og K antall kommuner i kontrollgruppen.

Dekomponeringen av tidsrekkene i disse faktorene er en rent maskinmessig måte å finne felles variasjonsmønstre $f_t^{(j)}$ i dataene på tvers av kommunene. De er analoge til prinsippkomponentene i PCA. Når disse mønstrene er funnet, har vi en prediksjon for en gitt kommune k bare ved å angi et antall J «factor loadings» $\lambda_k^{(j)}$. Hvis $J \ll T$, er det oppnådd betydelig kompresjon av mengden data vi trenger for å beskrive utviklingen i hver kommune.

Man har nå etablert en prediksjon både for kontrollgruppen og eksperimentgruppen, der $\hat{b}_{k,t} = Y_{k,t} - \hat{Y}_{k,t}^{(0)}$ er estimatet av intervensjonseffekten når k tilsvarer en nedbrent kommune. BMMS plotter i figur 5 og 6 de observerte data $Y_t = \langle Y_{k,t} \rangle$, der vinkelparentesene betyr midling over kommunene i eksperimentgruppen, og de kontrafaktiske prediksjonene $Y_t^{(0)} = \langle Y_{k,t}^{(0)} \rangle$. Avstanden mellom disse to kurvene, $\hat{b}_t = Y_t - Y_t^{(0)}$, er altså et mål for den instantane intervensjonseffekten. Verdien «ATT», som er angitt i figurene og i tabell A1, er verdien av \hat{b}_t tidsmidlet over alle valg etter krigen. Denne midlete effekten blir et ganske

grovt og upresist mål fordi effekten kan være positiv i noen perioder og negativ i andre. Det er mer informativt å studere figurene.

GSC uten kovariater kan skape spuriøse sammenhenger

BMMS gjør altså ingen forsøk på å introdusere kovariater, selv om både DiD og GSC åpner for slike generaliseringer. Å plukke ut én eksogen variabel og utelate andre plausible prediktorer regnes vanligvis som en oppskrift på deteksjon av en spuriøs sammenheng. Faktorene $f_t^{(j)}$ identifiseres utelukkende fra kontrollgruppens data. Det er altså ingen kjente årsaks-mekanismer som ligger til grunn, bare data, og de har ingen relasjon til intervensjonen. For oss som ikke driver med økonometri til daglig, kan dette virke nokså mystisk.

En antagelse, som i liten grad blir problematisert av Xu (2017), er at det er de samme underliggende uobserverte faktorene $f_t^{(j)}$ som styrer utviklingen av responsvariabelen i både eksperimentgruppen og kontrollgruppen, og at det bare er styrken av responsen (factor loading) som tillates å variere mellom kommuner. Argumentet for denne antagelsen synes å være at hvis de estimerte faktorene $f_t^{(j)}$ for kontrollgruppen viser seg å forklare en stor del av variansen i eksperimentgruppen før intervensjonen, taler dette for at de samme faktorene vil gi grunnlag for en god modell for disse kommunene i resten av tidsforløpet dersom nedbrenningen ikke hadde funnet sted.

Dette argumentet ser bort fra at krigen selv, og samfunnsutviklingen etter krigen, kan ha hatt ulik virkning på valgferden i eksperimentkommunene og i kontrollkommunene, uavhengig av nedbrenning. Slike forskjeller ville gi seg utslag et annet sett av faktorer $f_t^{(j)}$ enn de som er estimert fra kontrollkommunene, slik at GSC gir et uriktig estimat av det kontrafaktiske forløpet.

En metode for å undersøke om dette kan forekomme, er å konstruere placebo eksperimentgrupper fra den delen av kommunepopulasjonen som ikke er nedbrent. BMMS har allerede vist ett eksempel på dette i figur 6 ved å bruke *nord* som eksperimentgruppe og *sør* som kontroll. I perioden etter 1970 er \hat{b}_t like stor for *nord* vs. *sør* som *brent* vs. *nord*, mens et feilfritt GCS-estimat for *nord* vs. *sør* burde gi $\hat{b}_t = 0$, siden ingen kommuner i *nord* er brent. Når man tar i betraktning de store ulikhetene mellom Nord-Troms/Finnmark og resten av Nord-Norge, er det like gode grunner til at tilsvarende store spuriøse \hat{b}_t vil kunne opptre for *brent* vs. *nord*. Med andre ord indikerer dette at usikkerheten i \hat{b}_t -estimatet som skyldes antagelsene som CGS-metoden baserer seg på, er minst like stor som effekten selv, og det betyr at man ikke kan forkaste null-hypotesen om at nedbrenningen ikke har noen effekt på valgresultatene. Denne konklusjonen styrkes av det faktum at den påståtte effekten først viser seg flere tiår etter krigen, siden flere faktorer vil påvirke valgresultat jo lenger tid som går etter intervensjonen. Hypotesen om at effekten skyldes minner fra nedbrenningen som gir utslag først en generasjon senere, framstår som ad hoc.

Integrasjon av dynamisk beskrivelse i kausal design

Fra figur 2 i BMMS (og figur 2 her) ser man at trenden i alle grupper var økende før krigen, og flatet ut mellom 1945 og 1970 for deretter å avta fram mot 2017. Det mest dramatiske er jo utviklingen som skjedde i alle grupper. Den noe kraftigere nedgangen for den nedbrente gruppen etter 1970 ser ut som en liten perturbasjon i det store bildet av forandring, og det får meg til å stille spørsmålet om man virkelig kan forstå denne perturbasjonen bare som en respons på et eksogent sjokk uten å relatere det til det store bildet. Geografi, næringsliv, demografi og etnisk sammensetning var ulike i gruppene *brent*, *nord* og *sør* både før og etter

krigen. Likedan var det med oppslutningen om AP, som systematisk var større jo lenger nord man kom helt fra 1918 til 1970, og som deretter gradvis konvergente fram mot årtusenskiftet. Ved bare å betrakte disse rådataene er min første innskyttelse at de tre kommune-gruppene alle har økt oppslutningen om AP gjennom mellomkrigstiden, men at denne responsen har vært sterkere jo lenger nord man kommer. Utviklingen i etterkrigstiden har vært en stabilisering på førkrignivået fram til 1970, etterfulgt av en relaksasjon mot en felles likevekt.

Denne utviklingen kan modelleres uten å trekke inn nedbrenningen. Det viktigste sjokket som rammet Norge på 40-tallet, var en verdenskrig som tok slutt, og en ny verden med eventyrlig teknologisk utvikling og økonomisk vekst vokste fram. Ikke minst ble geografiske forskjeller stadig mindre betydningsfulle. Kunne denne geografiske homogeniseringen vært tilstrekkelig til å forklare den konvergensen i stemmetall mellom de tre gruppene som ses i figur 2, dersom den hadde vært inkludert som en kontrollvariabel? Det faktum at denne konvergensen også skjer mellom *nord* og *sør*, og altså uavhengig av nedbrenning, støtter denne hypotesen.

Lineære regresjonsmodeller og deres slektninger som DiD og GSC har til felles at de forklarer observasjoner ved å addere opp responsen til observerte prediktorer og uobserverte faktorer. Slike lineære modeller beskriver ikke alltid komplekse systemer som gjennomgår transformasjoner, og da er det et alternativ å ta utgangspunkt i en dynamisk modell av fenomenet før vi estimerer modellparametere fra data. Hypotesen om konvergens mot en likevekt som er identisk for alle norske kommuner, kan beskrives dynamisk, som jeg skal forsøke å illustrere med en analogi fra fysikken.

La oss tenke oss en kommune i mellomkrigstiden som en kjel med vann som varmes opp fra romtemperatur på en kokeplate. De ulike kommunene er bare forskjellige ved at kjelene inneholder ulik mengde vann. De settes på identiske kokeplater med identisk effekt og varmes opp i like lang tid. Så kommer krigen, platene slås på sparebluss i en periode og deretter helt av, og vannet i kjelene avkjøles og konvergerer mot rommets temperatur. Mengden vann i kjelene er ikke tilfeldig fordelt; den avtar jo lenger nord man kommer, og er aller lavest i Nord-Troms og Finnmark. Derfor er temperaturen høyest i disse kommunene når krigen kommer, lavere i *nord* og aller lavest i *sør*. Tidsforløpet av temperaturen i en gitt kjel kan beregnes ved hjelp av termodynamiske lover og gis en matematisk form som avhenger av én parameter som er ulik for hver kjel, nemlig vannmengden, som også bestemmer hvor høy temperaturen er når platen slås av. Hvis vi har observasjoner av tidsforløpet av temperaturen i en gitt kjel, kan vi bestemme denne parameteren ved hjelp av lineær eller ikke-lineær regresjon.

Jeg tror det vil være mulig å formulere en matematisk modell som ligner på vannkjelmodellen, og som vil forklare mye av tidsforløpene som er plottet i figur 2 i BMMS. Og dette er helt uten å innføre nedbrenningen som en forklaringsvariabel. En dynamisk beskrivelse kan stå på egne ben som en forklaringsmodell uten å involvere nedbrenning, men den kan også integreres i prediktorene (kovariatene) i regresjonsmodeller. Et eksempel fra klimaforskning er beskrevet i Rypdal (2018), der observerbare klimapådriv «filtreres» til anvendbare prediktorer for global overflatetemperatur ved å bruke kunnskap vi har om klimasystemets dynamiske impulsrespons på slike pådriv. Innføring av en slik prediktor som ekstra kovariat i BMMS' modell kunne muligens avgjøre hvilken av de to forklaringene, nedbrenning eller relaksasjon til likevekt, som er den mest sannsynlige.

Konklusjon

Jeg har argumentert for at det kan være gode grunner til å sette spørsmålstegn ved resultater som i hovedsak bygger på avanserte økonomiske metoder i samfunnsfagene, og spesielt hvis de følger av rutinemessig bruk av ferdige programpakker slik som den som er utviklet av Xu og Liu (2022). Likevel er det ingen tvil om at slangen er kommet til paradiset for å bli. Med rasjonell skepsis og sunn fornuft kan disse metodene bli et verdifullt tilskudd og er bedre enn både total avvisning og kritikkløs tilbedelse av kvantitative metoder som er vanskelige tilgjengelige for de fleste av oss.

Referanser

- Abadie, A. (2021). Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects. *Journal of Economic Literature* 59(2), 391–425. <https://doi.org/10.1257/jel.20191450>
- Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust difference-in-differences estimates? *The Quarterly Journal of Economics*, 119(1), 249–275. <https://doi.org/10.1162/003355304772839588>
- Buck, M., Midtbø, S., Markussen, J.A., & Stein, J. (2022). Brent jord: langsiktige politiske konsekvenser av nedbrenning og tvangsevakuering i Nord-Troms og Finnmark. *Tidsskrift for samfunnsforskning*, 63(2), 79–101. <https://doi.org/10.18261/tfs.63.2.1>
- Rypdal, K. (2018). The Life and Death of the Recent Global Surface Warming Hiatus Parsimoniously Explained. *Climate*, 6(3), 64. <https://www.mdpi.com/2225-1154/6/3/64>
- Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*, 25(1), 57–76. <https://doi.org/10.1017/pan.2016.2>
- Xu, Y., & Liu, L. (2022). Package 'gsynth'. <https://cran.r-project.org/web/packages/gsynth/gsynth.pdf>