

Observationsstudier II

Erik Gahner Larsen

Kausalanalyse i offentlig politik

Eksamensdato

- ▶ Dato for aflevering: 2. juni, klokken 12.00
- ▶ Nye datoer
 - ▶ Tentativ udgave: 15. maj
- ▶ Vejledning
 - ▶ Ideelle tidspunkter?

Gæsteforelæsning

- ▶ Evalueringen af offentlige politikker
- ▶ Mandag den 18. april
- ▶ Se den opdaterede læseplan!

- ▶ Tid: Onsdag d. 4. maj
 - ▶ Hele dagen fra kl. 08.45 (al undervisning er aflyst)
- ▶ Sted: SDU O100 - ved hovedindgangen
- ▶ Følgende oplægsholdere er allerede præsenteret (flere på vej):
 - ▶ Martin Lidegaard, medlem af folketinget for Radikale Venstre og forhenværende udenrigsminister
 - ▶ Flemming Rose, fhv. kultureddaktør på Jyllands-Posten
 - ▶ Andreas Kamm, generalsekretær i Dansk Flygtningehjælp
- ▶ Yderligere oplysninger på
<https://www.facebook.com/AGORASDU/>

Dagsorden

- ▶ Observationsstudier, recap
- ▶ Mange observationer: Matching metoden
- ▶ Færre (aggregerede) observationer: Syntetisk kontrol metoden
 - ▶ Komparative casestudier

Observationsstudier

- ▶ Husk: Observationsstudier er studier hvor forskeren...
 - ▶ ikke kontrollerer datagenerationsprocessen og
 - ▶ der ikke er randomiseret variation
- ▶ Logikken i eksperimenter og observationsstudier er den samme
 - ▶ Problem: Sammenligningen af observationer, der ikke er sammenlignelige
 - ▶ Stimulitildelingen er ikke tilfældig
- ▶ Vi skal approksimere det eksperimentelle ideal i vores observationsstudier

Observationsstudier og kausal inferens

- ▶ Vi har stadig et stimuli!
 - ▶ “A study without a treatment is neither an experiment nor an observational study.” (Rosenbaum 2002, 1)
- ▶ To grupper: Stimuli og kontrol
- ▶ I eksperimenter:
 - ▶ Grupperne er sammenlignelige (med undtagelse af stimuli)
 - ▶ Kontrolvariable påvirker ikke stimultildeling
 - ▶ Sandsynligheden for at få stimuli er kendt
 - ▶ På grund af randomisering
 - ▶ $P(W_i) = 0.5$, for alle i ved stimuli og kontrol
 - ▶ Uafhængig: $(Y(1), Y(0), X) \perp W$

Observationsstudier og kausal inferens

- ▶ De bedste observationsstudier bygger på logikken fra eksperimenter
- ▶ I eksperimenter er vi interesseret i gennemsnitseffekter (ATE)
- ▶ I mange observationsstudier er vi interesseret i gennemsnitlige stimulieffekter for stimuligruppen (ATT)
 - ▶ $ATT = E[Y(1) - Y(0)|W = 1]$
- ▶ Giver os mulighed for at evaluere effekten for de enheder der *har* fået stimuli
- ▶ Kontrafaktisk gennemsnit: $E[Y(0)|W = 1]$

Observationsstudier og kausal inferens

- ▶ I observationsstudier kender vi ikke stimultildelingsmekanismen
 - ▶ Vi ved blot, at den ikke altid er tilfældig
 - ▶ $P(W_i) \neq 0.5$
- ▶ Når kontrolvariable, X , påvirker stimultildelingen: Matching
 - ▶ “Matching refers to a variety of procedures that restrict and reorganize the original sample in preparation for a statistical analysis.” (Gelman og Hill 2007, 206)

Matching

- ▶ Med andre ord: Matching hjælper os med at skabe mere sammenlignelige grupper
- ▶ Hvordan?
 - ▶ Maksimerer balancen på kontrolvariable
- ▶ Gør stimuli og kontrolgruppe identitiske før stimultildeling
- ▶ Fundamental antagelse: balance på observerede variable → balance på uobserverede variable

Matching

- ▶ Resultat: Vi kan sammenligne sammenlignelige enheder
 - ▶ Hvad kan vi lære fra enheder, vi ikke kan sammenligne med noget?
 - ▶ Intet. Mangler kontrafaktisk scenarie.
- ▶ Løsningen er et distancemål (D_{ij}), der måler distancen mellem to enheder i forhold til deres værdier på kontrolvariablene
- ▶ Der findes forskellige distancemål

Eksakt matching (stratificeret matching)

- ▶ To muligheder for afstand mellem to enheder
 - ▶ Ingen afstand mellem observationer
 - ▶ Uendelig afstand

$$D_{ij} = \begin{cases} 0 & \text{if } X_i = X_j \\ \infty & \text{if } X_i \neq X_j \end{cases}$$

- ▶ Problem: Dimensionalitetsproblemet / *curse of dimensionality* (Sekhon 2009, 497)
- ▶ Fungerer bedst med få variable der er kategoriske

Propensity score

- ▶ En løsning på dimensionalitetsproblemet
- ▶ Propensity score: “the propensity towards exposure to treatment 1 given the observed covariates x ” (Rosenbaum og Rubin 1983, 43)
- ▶ Propensity score: $p_i \equiv Pr(W_i|X_i)$
- ▶ Sandsynlighed for at få stimuli givet kontrolvariable
- ▶ Distance: $D_{ij} = |p_i - p_j|$
- ▶ I praksis: En propensity score for hver enhed (en ekstra kolonne i vores datasæt)
 - ▶ Evt. de forudsagte sandsynligheder fra en logistisk regression

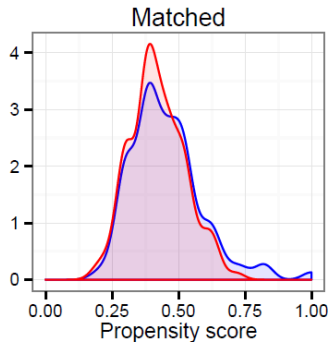
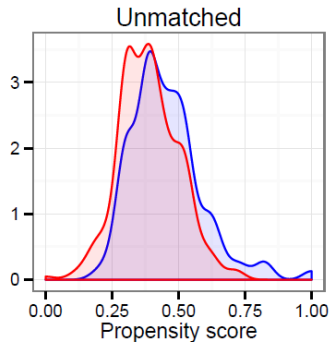
Designantagelser

- ▶ Antagelse 1: $Pr[W|X, Y(1), Y(0)] = Pr(W|X)$
(*Unconfoundedness*)
- ▶ Forskellige personer har forskellige *propensity scores* (Rubin 2004). Eksempler:
 - ▶ Ældre mænd har sandsynligheden 0,8 for at få stimuli
 - ▶ Yngre mænd har sandsynligheden 0,6 for at få stimuli
 - ▶ Ældre kvinder har sandsynligheden 0,5 for at få stimuli
 - ▶ Yngre kvinder har sandsynligheden 0,2 for at få stimuli

Designantagelser

- ▶ Antagelse 2: $0 < p_i < 1$ (sandsynligheden er mellem 0 og 1, i.e. overlap)
- ▶ Overlap (*common support*)
- ▶ Hvad hvis $p_i = 1$ eller $p_i = 0$?
- ▶ Ekskluder observationer for hvem p_i er tæt på 0 eller 1 (tommelfingerregel: ekskluder ved $p_i < 0.1$ og $p_i > 0.9$)

Eksempel, overlap før og efter matching



Overlap før og efter matching

- ▶ Tjek om der er overlap før og efter
- ▶ Har vi fået identiske grupper?
 - ▶ I så fald: ingen observerbare forskelle

Hvad er et godt match?

- ▶ Hvor tætte skal to observationer være, før vi har et godt match?
- ▶ Vi smider observationer ud fra vores datasæt
- ▶ Sæt en *caliper* der er grænseværdien for, hvor stor en distance vi vil acceptere
- ▶ Pre-stimuli kontrolvariable
 - ▶ Vælg relevante kontrolvariable
 - ▶ Relevante
 - ▶ Pre-stimuli (*Ikke* noget, der kan blive påvirket af stimuli)

Matching, forskellige metoder

- ▶ Nearest neighbor matching (med eller uden caliper)
- ▶ Radius matching
- ▶ Genetic matching
- ▶ Coarsened exact matching

Nearest neighbor og specifikke valg

- ▶ Nærmeste nabo. Vælg den *nabo* i kontrolgruppen, der ligger tættest på stimulienheden
- ▶ Trade-off: Bias og varians
- ▶ Antal matches
 - ▶ Match 1 NN: Mindre bias, mere varians
 - ▶ Match 1:n NN: Mindre varians, mere bias
- ▶ Med eller uden erstatning?
 - ▶ Med erstatning: Mindre bias, mere varians
 - ▶ Uden erstatning: Mindre varians, potentiel bias
 - ▶ Tommelfingerregel: match med erstatning

Balance

- ▶ Forskellige måder at teste balancen
- ▶ Forskelle i andele/gennemsnit (t-test, Fisher exact test)
- ▶ Distributionen (QQ plot, Kolmogorov-Smirnov test)
- ▶ Igen: Vi ønsker identitiske grupper (med undtagelse af stimuli)
- ▶ Ingen observerbare forskelle

Hvorfor brugere forskere matching? (Miller 2015, 31)

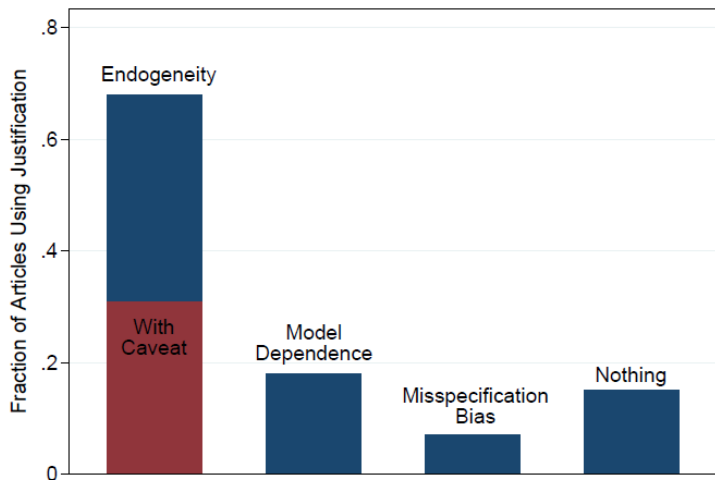
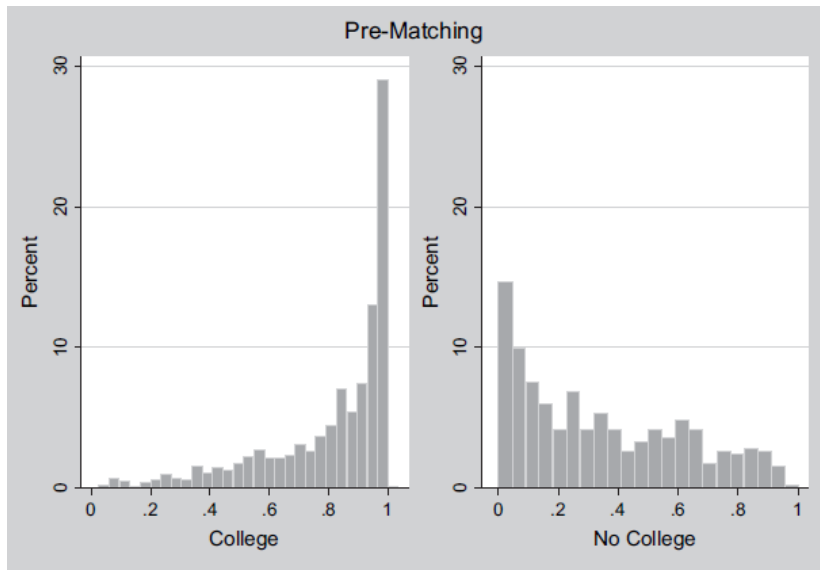


Fig. 1: Based on a survey of 68 articles using matching published since 2009, the figure shows the fraction of articles that reference each of the three justifications for matching analyzed in this article, as well as the fraction that offer no justification.

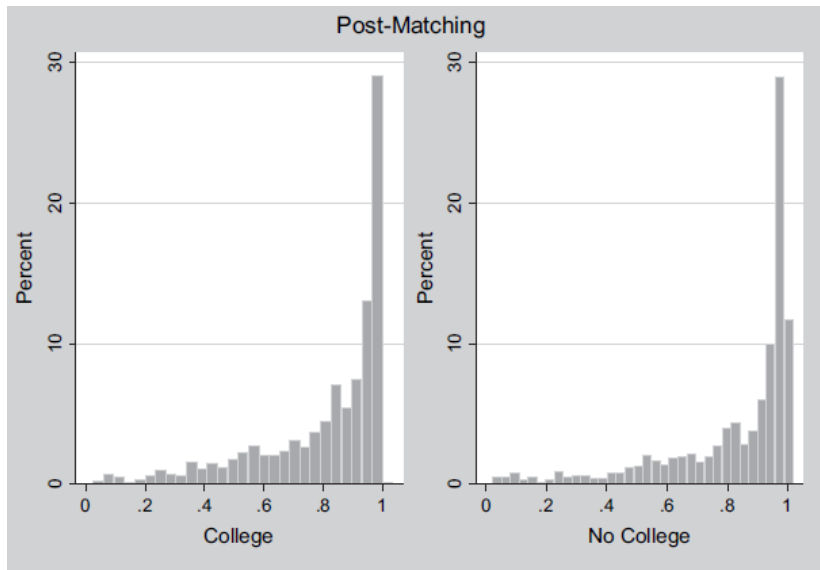
Eksempel: Den kausale effekt af uddannelse på politisk deltagelse

- ▶ Kam og Palmer (2008): Et klassisk stimultildelingsproblem
- ▶ Logikken: “matches respondents who attended college with those who did not by using a propensity score, or predicted likelihood of attending college based upon an individual’s preadult experiences and characteristics. The matching process mimics random assignment, thus producing two groups whose levels of participation can then be compared, having essentially controlled for preadult experiences and characteristics.” (p. 613)
- ▶ “Each respondent who received the treatment (i.e., a respondent who went to college, in our research question) is matched with a set of untreated respondents (i.e., respondents who did not attend college) that have similar propensity scores. This technique is called nearest-neighbor matching.” (Kam og Palmer 2008, 620)

Propensity scores, før matching



Propensity scores, after matching



Eksempel: Den kausale effekt af uddannelse på politisk deltagelse

- ▶ Problem: Propensity scores tæt på 1
- ▶ Henderson og Chatfield (2011, 652): “we observe that clustering around 1 is so pronounced that the p-scores for the top 5% of college-attenders range in value from .9998174 to .9999998. In fact, over half of all attenders have propensity scores greater than .9 and over a quarter have scores greater than .99. In contrast, only 14 nonattenders have propensities greater than .9 (3%) and only 5 (1%) have propensities greater than .95, with the highest propensity for any nonattender being .9889.”

Eksempel: Den kausale effekt af uddannelse på politisk deltagelse

- ▶ Kam og Palmer (2008): Ingen effekt
- ▶ Mayer (2011): Uddannelse øger den politiske deltagelse
- ▶ Henderson og Chatfield (2011): Uddannelse øger den politiske deltagelse
- ▶ Kam og Palmer (2011): Ingen kausal effekt

Eksempel: Effekten af samfundstjeneste på kriminalitet

- ▶ Klement (2015): Comparing the effects of community service and imprisonment on reconviction: results from a quasi-experimental Danish study
- ▶ Udfaldsvariabel: Domsrate
- ▶ Stimuli: Samfundstjeneste (kontrolgruppe: fængsel)
- ▶ Design: Kvasiekperiment
- ▶ Sample: Danske kriminelle
- ▶ Resultat: Samfundstjeneste → Mindre tilbagefald

Unmatched

	Unmatched data		
	CS (n= 1021)	Prison (n=604)	Sig.
<i>Attributes</i>	%	%	
<i>Index crime type</i>			***
Crimes against the person (1)	49	70	
Property crime (4)	38	14	
Drug-related crimes (5)	8	10	
Other types of crime (6)	5	6	

* $p < .05$; ** $p < .01$; *** $p < .001$

Matched

	Unmatched data			Matched data		
	CS (n=1021)	Prison (n=604)	Sig.	CS (n=998)	Prison (n=604)	Sig.
<i>Attributes</i>	%	%		%	%	
<i>Index crime type</i>			***			***
Crimes against the person (1)	49	70		50	70	
Property crime (4)	38	14		37	14	
Drug-related crimes (5)	8	10		8	10	
Other types of crime (6)	5	6		5	6	

* $p < .05$; ** $p < .01$; *** $p < .001$

Begrænsninger og udfordringer med matching

- ▶ Intet studie er bedre end dets design
 - ▶ “Without an experiment, a natural experiment, a discontinuity, or some other strong design, no amount of econometric or statistical modeling can make the move from correlation to causation persuasive.” (Sekhon 2009, 503)
 - ▶ “Matching has no advantage relative to regression for inferring causation or dealing with endogeneity” (Miller 2015, 2)
- ▶ Flere trin, flere steder det kan gå galt
 - ▶ ‘Researcher Degrees of Freedom’
 - ▶ Misvisende forskning
 - ▶ Først når vi har sammenlignelige grupper, bør man estimere den kausale effekt

Begrænsninger og udfordringer med matching

- ▶ Matching løser ikke FPCI
- ▶ Gode designs er løsningen
 - ▶ Randomisering \rightarrow balance
 - ▶ Balance \nrightarrow randomisering
- ▶ Hovedpointe: viden om observerbare forskelle mellem stimuli og kontrolgrupper kan hjælpe os med at reducere bias i vores effektestimater

Syntetisk kontrol

- ▶ Ved færre observationer: syntetisk kontrol metoden for komparative case studier
- ▶ Målsætning: Estimere effekten af stimuli/intervention på få observationer
- ▶ Især applicerbart på komparative casestudier
 - ▶ Ét land gennemfører en reform, andre gør ikke
 - ▶ Én stat i USA forbyder noget, andre gør ikke
- ▶ Største svaghed: arbitrært valg af kontrolgruppe
- ▶ Vi er interesseret i et vægtet gennemsnit af alle potentielle sammenligningsenheder der bedst ligner den enhed, vi ønsker at studere
 - ▶ Ligesom med matching

Komparative casesstudier

- ▶ Design: Afbrudt tidsseriedesign med en kontrolgruppe
 - ▶ En eller flere målinger før og efter intervention
 - ▶ En stimuligruppe/enhed får intervention
 - ▶ En kontrolgruppe/enhed får ikke intervention
- ▶ Hvordan laver vi de rette kontrafaktiske sammenligninger?
- ▶ Hvilke enheder skal vi sammenligne?
 - ▶ To eller flere?

Syntetisk kontrol

- ▶ Reform finder sted i Danmark, men ikke i andre lande
 - ▶ Vi har værdier før og efter reformen i Danmark og andre lande
 - ▶ Det vi ønsker at lave er en kontrolgruppe
 - ▶ Et syntetisk Danmark
- ▶ Begivenhed finder sted i Odense Kommune, men ikke andre kommuner
 - ▶ Svært at sammenligne Odense Kommune med de andre kommuner
 - ▶ Samme opgave: Lave en syntetisk kontrol
- ▶ Syntetisk kontrol
 - ▶ Statistisk værktøj til kvalitative studier

Model

- ▶ $J+1$ regioner, T perioder
 - ▶ J : mulige kontrolregioner
 - ▶ T : antal perioder
 - ▶ T_0 : antal prestimuli perioder ($1 \leq T_0 < T$)
- ▶ Region 1 påvirkes af stimuli i perioden $T_0 + 1, \dots, T$.
- ▶ Y_{jt}^N : Udfald for region j i periode t uden stimuli/intervention
- ▶ Y_{jt}^I : Udfald for region j i periode t med stimuli/intervention
- ▶ Effekt af stimuli/intervention: $\alpha_{jt} = Y_{jt}^I - Y_{jt}^N$

Model

- ▶ Intervention, D_{jt} : 1 hvis region j er tildelt stimuli i periode t
- ▶ Observeret udfald for region j i periode t :

$$Y_{jt} = Y_{jt}^N + \alpha_{jt} D_{jt}$$

- ▶ Kun region 1 får stimuli - og kun efter T_0

$$D_{jt} = \begin{cases} 1 & \text{if } j = 1 \text{ and } t > T_0 \\ 0 & \text{otherwise.} \end{cases}$$

- ▶ Hovedinteresse: Estimere α_{1t} , altså effekten af interventionen på region 1
 - ▶ $\alpha_{1t} = Y_{1t} - Y_{1t}^N$
 - ▶ Husk: Y_{1t} er observeret
 - ▶ Vi er interesseret i estimationen af det kontrafaktiske scenarie

Model

- ▶ Kontrafaktisk scenarie er givet ved en faktormodel:
 - ▶ $Y_{jt}^N = \delta_t + \theta_t \mathbf{Z}_j + \lambda_t \mu_j + \varepsilon_{jt}$
 - ▶ Hvor:
 - ▶ δ_t (delta): ukendt faktor med konstant faktorladning på tværs af enheder
 - ▶ θ_t (theta): ukendte parametre
 - ▶ \mathbf{Z}_j : kontrolvariable der ikke er påvirket af stimuli
 - ▶ λ_t (lambda): ukendte fælles faktorer
 - ▶ μ_j (my): ukendte faktorladninger
 - ▶ ε (epsilon): uobserverede kortvarige chok
 - ▶ Hovedpointe: Forskellige reaktioner til flere uobserverede faktorer

Model

- ▶ Uobserverede faktorer kan være problematiske
 - ▶ Vi kan matche på pre-stimuli perioderne for at kontrollere for heterogene reaktioner til uobserverede faktorer
- ▶ Potentiel syntetisk kontrol givet et sæt af vægte, W
 - ▶ $W = (w_2, \dots, w_{J+1})'$
 - ▶ Hver vægt er større end nul ($w_j \geq 0$ for $j = 2, \dots, J + 1$)
 - ▶ Summen af disse værdier er 1 ($w_2 + \dots + w_{J+1} = 1$)
 - ▶ For hver værdi er der en syntetisk kontrol
 - ▶ Et vægtet gennemsnit for kontrolregionerne
- ▶ Udfaldet for hver syntetisk kontrol er givet ved:

$$\sum_{j=2}^{J+1} w_j Y_{jt} = \delta_t + \theta_t \sum_{j=2}^{J+1} w_j \mathbf{Z}_i + \lambda_t \sum_{j=2}^{J+1} w_j \mu_j + \sum_{j=2}^{J+1} w_j \varepsilon_{jt}$$

Model

- ▶ Den syntetiske kontrol estimator er så:

$$\hat{\alpha}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j Y_{jt}$$

- ▶ Y_{jt} : udfald for region j i periode t .
- ▶ $\hat{\alpha}_{1t}$: effekten af intervention/stimuli
- ▶ Kontrafaktisk scenarie: kombinationen af enheder, der ikke har fået stimuli, som bedst ligner vores interventionsenhed før interventionen
- ▶ Tænk en tidsseriegraf hvor vi vil have kontrol og stimuli til at være identiske før en intervention

Eksempel: Californiens Proposition 99 (Abadie et al. 2010)

- ▶ Politik: Øgede priser på cigaretter (øget skat)
- ▶ Data: Paneldata for amerikanske stater (1970-2000)
- ▶ Udfald: Tobaksforbrug
- ▶ Interventionsår: 1988
- ▶ Region: Californien
- ▶ Syntetisk Californien: vægtet gennemsnit af potentielle kontrolstater
 - ▶ Vægte valgt med henblik på at den syntetiske kontrol bedst forudsiger værdierne på predictors af cigaretforbrug før interventionen (*Proposition 99*)
- ▶ Andre stater der efterfølgende implementerede store kontrolprogrammer er ekskluderet

Eksempel: Californiens Proposition 99 (Abadie et al. 2010)

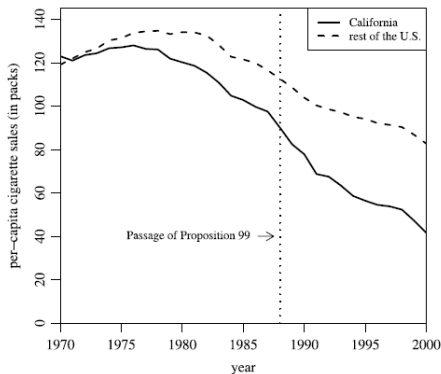


Figure 1. Trends in per-capita cigarette sales: California vs. the rest of the United States.

Eksempel: Californiens Proposition 99 (Abadie et al. 2010)

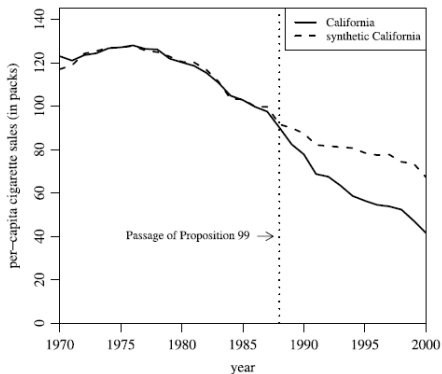


Figure 2. Trends in per-capita cigarette sales: California vs. synthetic California.

Eksempel: Californiens Proposition 99 (Abadie et al. 2010)

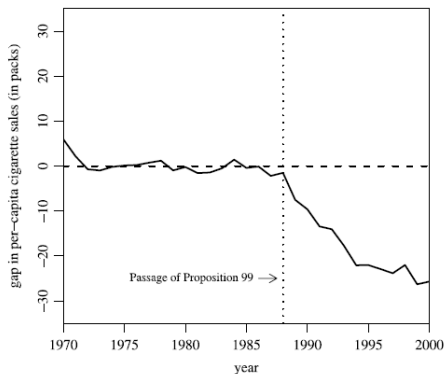


Figure 3. Per-capita cigarette sales gap between California and synthetic California.

Eksempel: Californiens Proposition 99 (Abadie et al. 2010)

- ▶ Placebotests
- ▶ Hvordan?
 - ▶ Interventionen bliver givet til hver af de 38 kontrolstater
 - ▶ Estimerer stimuleffekten for hver stat, hvis den implementerede kontrolprogrammet i stedet for Californien
- ▶ RMSE

Eksempel: Californiens Proposition 99 (Abadie et al. 2010)

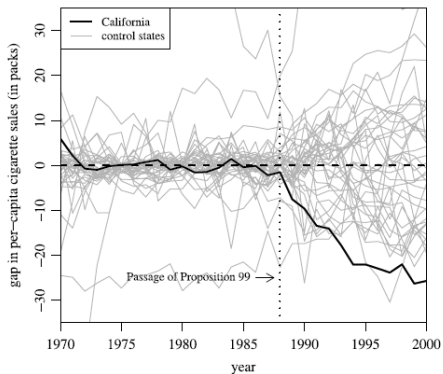


Figure 4. Per-capita cigarette sales gaps in California and placebo gaps in all 38 control states.

Eksempel: Californiens Proposition 99 (Abadie et al. 2010)

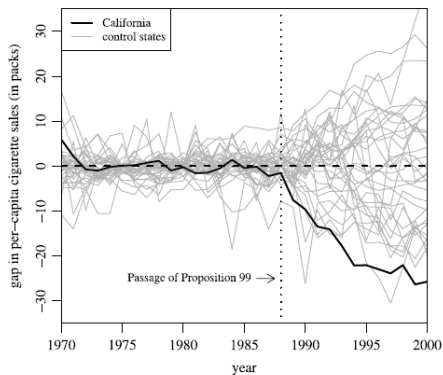


Figure 5. Per-capita cigarette sales gaps in California and placebo gaps in 34 control states (discards states with pre-Proposition 99 MSPE twenty times higher than California's).

Eksempel: Effekten af terror på økonomien (Abadie og Gardeazabal 2003)

- ▶ Region: Baskerlandet
- ▶ Udfald: BNP pr. indbygger
- ▶ Intervention: Terror
- ▶ “We approach this problem by comparing the economic evolution of the Basque Country during the terrorist era with that of a weighted combination of other Spanish regions chosen to resemble the characteristics of the Basque Country before terrorism. We conceptualize such a weighted average of other Spanish regions as a “synthetic” Basque Country without terrorism, against which we can compare the actual Basque Country with terrorism.” (p. 116)

Eksempel: Effekten af terror på økonomien (Abadie og Gardeazabal 2003)

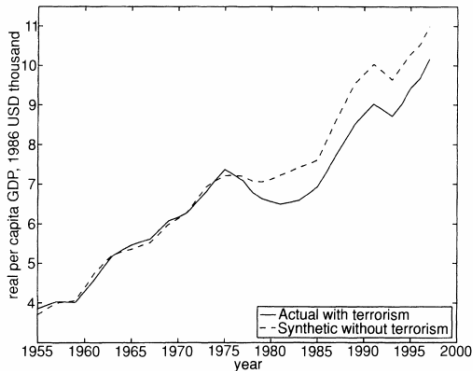


FIGURE 1. PER CAPITA GDP FOR THE BASQUE COUNTRY

Eksempel: Effekten af Tysklands genforening på Vesttysklands økonomi (Abadie et al. 2015)

FIGURE 1 Trends in per Capita GDP: West Germany versus Rest of the OECD Sample

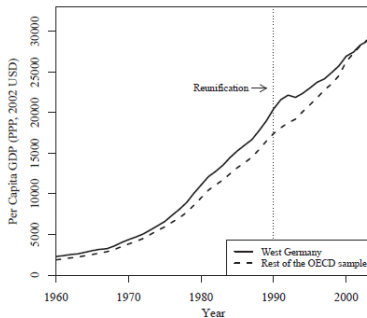
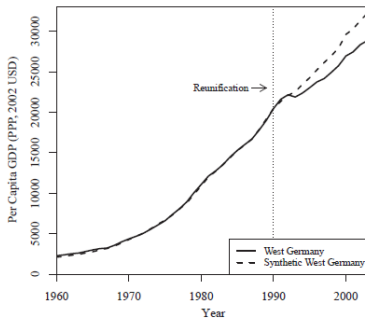
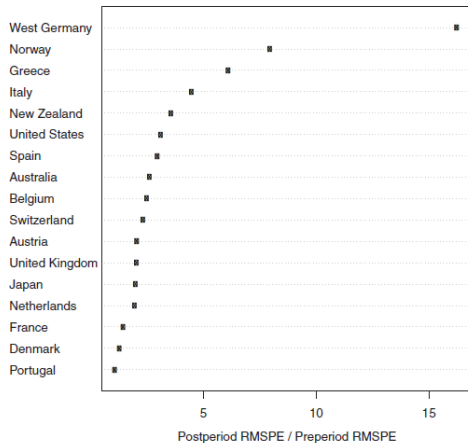


FIGURE 2 Trends in per Capita GDP: West Germany versus Synthetic West Germany

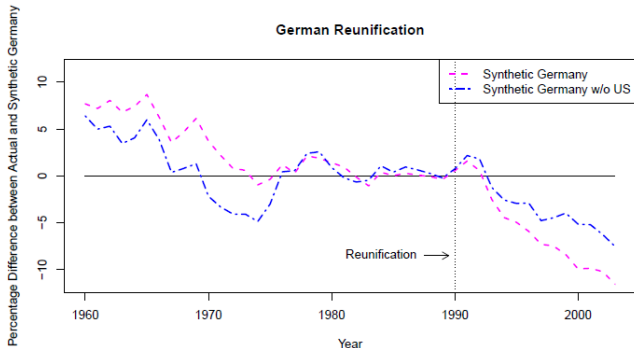


Eksempel: Effekten af Tysklands genforening på Vesttysklands økonomi (Abadie et al. 2015)

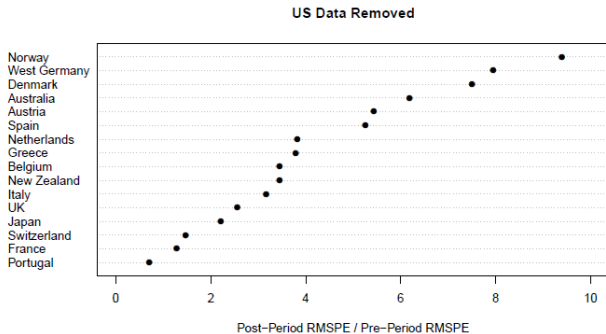
FIGURE 5 Ratio of Postreunification RMSPE to Preunification RMSPE: West Germany and Control Countries



Eksempel: Effekten af Tysklands genforening på Vesttysklands økonomi (Klößner et al. 2016)



Eksempel: Effekten af Tysklands genforening på Vesttysklands økonomi (Klößner et al. 2016)



Hvad med SUTVA?

- ▶ Vi antager stadig, at der ikke er interaktioner mellem enhederne
- ▶ Hvorfor? Ikke muligt at estimere effekten af en intervention hvis der er brud på SUTVA

Matching og syntetisk kontrol i Stata og R

- ▶ Matching: forskellige pakker i både Stata og R
 - ▶ Kontakt mig hvis I har brug for hjælp
- ▶ Syntetisk kontrol: brug Synth pakken til både Stata og R
 - ▶ Kontakt mig hvis I har brug for hjælp

Næste gang

- ▶ Gæsteforelæsning
- ▶ Mandag, kl. 10-12 i U31A