

Asmus Leth Olsen

# Tærskelvariable og tærskelværdier: en introduktion til regressions- diskontinuitetsdesignet<sup>1</sup>

Regressionsdiskontinuitetsdesignet (RDD) udnytter empiriske situationer, hvor vi kan observere en variabel, der ved en given værdi opdeler de observationer, vi studerer, i en kontrol- og forsøgsgruppe. RDD har vist sig at reproducere traditionelle, eksperimentelle resultater, være intuitivt enkelt og relativt nemt at implementere i statistiske programpakker. Alligevel er det først inden for de seneste par år, at RDD er blevet anvendt i statskundskaben. Oftest har krikken af designet været, at den empiriske virkelighed sjældent tilbyder situationer, hvor RDD er anvendeligt. Denne artikel introducerer forskere og studerende til RDD i en statskundskabssammenhæng. Der vil være et særligt fokus på at vise, at RDD faktisk er brugbart til at besvare ganske mange forskelligartede kausale spørgsmål i statskundskaben.

Observationsstudier er karakteriseret ved, at vi som forskere eller studerende ikke aktivt har inter文neret for at opdele de individer eller organisationer (herafter: observationer), vi studerer, i en kontrol- og forsøgsgruppe. Vi er derfor henvist til at studere observationer, der selv har selekteret sig ind i kontrol- og forsøgsgrupper eller er blevet selekteret af andre (Robinson et al., 2009). Idéen om kontrol- og forsøgsgrupper er således alene et udtryk for, at observationerne har forskellige værdier på den variabel, som vi ønsker at estimere den kausale effekt af. Hvorfor de er endt i den ene eller anden gruppe, er derfor ude af vores hænder. Denne (selv)selektion betyder, at observationerne formentlig er havnet i en særlig gruppe af gode grunde, hvorfor der som oftest vil være systematiske forskelle mellem kontrol- og forsøgsgruppe. Dette giver os et potentielt *confounding* problem, da der vil være en række faktorer, der påvirker både observationernes placering i kontrol- eller forsøgsgruppe og den afhængige variabel, vi vil undersøge. Meget sjældent har vi den fornødne viden om verden eller præcise nok data til at tage højde for confounding (Freedman, 1991). Men i en kort, simpel artikel foreslog Thistlethwaite og Campbell (1960) regressionsdiskontinuitetsdesignet (RDD), som et muligt design til at besvare kausale spørgsmål i tilfælde, hvor vi direkte kan observere selektionsprocessen i observationsstudier og dermed omgå problemet med confounding.

I korte træk udnytter RDD empiriske situationer, hvor vi kan observere en variabel, der ved en given værdi opdeler observationerne i en kontrol- og forsøgsgruppe. Der er altså et punkt på en variabel, som vi observerer, der slår en given intervention til og fra. Interventionen er som i andre observationsstudier stadig ikke vores værk eller udtryk for et formelt lodtrækningsforsøg, som i de traditionelle eksperimenter. RDD udnytter i stedet, at den præcise placering af fordelingen af observationer omkring punktet, hvor interventionen indtræder, kan være tilnærmelsesvis tilfældig (Dunning, 2012). Hermed kan vi estimere en kausal effekt af en intervention ved at sammenligne forskelle i gennemsnit på den afhængige variabel for observationer omkring interventionspunktet, hvor opdelingen i kontrol- og forsøgsgruppe sker. I RDD sker dette ved, at vi modellerer den variabel, der opdeler observationerne og/eller begrænser vores analyser til de observationer, der ligger tættest ved den værdi, som skiller kontrol- fra forsøgsgruppen (Green et al., 2009; Dunning, 2012). Den centrale antagelse er, at observationer tæt ved denne værdi ikke nøjagtigt har kunnet påvirke, om de er endt i kontrol- eller forsøgsgruppen, og at de ikke er blevet placeret intentionelt af andre.

Med afsæt i denne simple idé har RDD langsomt spredt sig på tværs af socialvidenskaberne (Cook, 2008). Og der er flere gode grund til, at RDD også bør vinde større indpas i statskundskaben: For det første kan RDD estimere kausale effekter, som kommer meget tæt på tilsvarende traditionelle, eksperimentelle kausale effekter (Aiken et al., 1998; Berk et al., 2010; Green et al., 2009; Shadish, 2011). For det andet er RDD intuitivt forståeligt, og det kræver relativt milde antagelser sammenlignet med andre naturlige eksperimentelle designs (Lee og Lemieux, 2009: 1; Dunning, 2012: 136). Sekhon (2009: 503) har argumenteret for, at observationsstudiers største udfordring i forhold til kausale spørgsmål er, at vi ikke har fundet en måde at masseproducere designs på i samme omfang som det traditionelle eksperiment lader sig masseproducere. Vi har altså ikke et simpelt *quick fix*, når vi arbejder med observationsdata. I stedet har forskellige former for regressionsanalyse, hvor vi på bedste vis kontrollerer for mulige *confounders*, været det typiske udgangspunkt for at estimere kausale effekter i observationsstudier i statskundskaben. Men som Freedman (1991) har argumenteret for, har denne praksis været en vildfarelse for statskundskaben, da vi ikke har det teoretiske fundament til at finde de rette kontrolvariable eller modellere dem med den korrekte funktionelle form. RDD er måske det tætteste, vi kommer på et sådant *quick fix* til kausal inferens med observationsdata: Et design der i sin idé er lige så simpelt som det traditionelle eksperiment (Morgan og Winship, 2007: 251), og som samtidig har vist sig at kunne producere lige så troværdige estimater af kausale effekter.

Alligevel har RDD allerede for længe kunnet fejre 50 års fødselsdag uden at have vundet bred udbredelse som design til at afdække kausale relationer i statskundskaben. Først inden for de seneste få år er en række RDD-studier blevet publiceret. En forklaring på den manglende udbredelse af RDD har været, at selvom designet er stærkt og intuitivt, er det i realiteten ret sjældent, at virkeligheden tilbyder muligheder for, at designet faktisk kan finde anvendelse (Cook, 2008; Lee og Lemieux, 2009). Eller rettere: RDD kræver, at forskeren afsøger empiriske områder, hvor designet kan finde anvendelse. Dvs. forskerens arbejde forskyder sig derfor fra statistisk modellering og over mod at søge empiriske kontekster, der tillader brug af RDD. I de seneste år har statskundskaben langsomt fået øjnene op for, at designet faktisk er relevant til at besvare kausale spørgsmål i alt fra valgforskning til offentlig politik (Lee, 2008; Olsen, 2013a).

I denne artikel vil jeg introducere RDD til forskere og studerende inden for statskundskaben. Hovedvægten vil være rettet mod at kommunikere idéen bag designet og kun i mere begrænset omfang selve estimationen, som er behandlet mere fyldestgørende af andre (Green et al., 2009; Dunning, 2012). Det primære formål vil være at stimulere til, at flere tænker over, hvordan RDD kan anvendes til at estimere kausale effekter inden for netop deres forskningsområde.

Artiklen er struktureret på følgende vis: Først præsenteres nogle af de grundlæggende begreber i RDD. Dernæst følger de centrale antagelser for at estimere kausale effekter med RDD og en introducerende beskrivelse af, hvordan RDD-modeller estimeres statistisk. Herefter kommer eksempler på brug af RDD inden for statskundskaben. Inden konklusionen præsenteres nogle barrierer for brugen af RDD i statskundskaben samt mulige løsninger.

## Regressionsdiskoninuitetsdesignet: de centrale begreber

Naturlige eksperimenter betegner empiriske forhold, hvor observationer tildeles værdier på en variabel af interesse på baggrund af tilnærmelsesvis tilfældige processer i "naturen" (Dunning, 2012; Robinson et al., 2009). Hermed opnår vi eksogen variation på den variabel, hvis kausale effekt vi ønsker at estimere uden aktiv intervention fra vores side. RDD skal her ses som en særlig klasse af naturlige eksperimenter, der sætter specifikke krav til de empiriske omstændigheder, herunder hvordan de undersøgte observationer har fået tildelt værdier på den variabel, som vi ønsker at estimere den kausale effekt af. Udgangspunktet i RDD er, at vi med deterministisk viden om selektionsprocessen kan tage højde for den og derved identificere en kausal effekt (Rubin, 1977; Shadish et al., 2001).

RDD består i korte træk af tre elementer: en tærskelvariabel, en tærskelværdi og en afhængig variabel, som vi ønsker at undersøge. I realiteten kan der godt

være flere tærskelvariable og tærskelværdier (Papay et al., 2011), men her ser vi alene på det simple eksempel med én tærskelvariabel og én tærskelværdi. Tærskelvariablen (*assignment*, *running* eller *forcing* variabel) er en kontinuerlig før-interventionsvariabel, som vi observerer for alle de enheder, vi undersøger (Shadish et al., 2001). Jo mere finkornet tærskelvariablen er desto bedre, da den skal give os så præcis som mulig information om, hvorvidt en observation modtager en intervention af interesse eller ej. Tærskelvariablen skal deterministisk bestemme, om observationer tilhører vores kontrol- eller forsøgsgruppe. Tærskelvariablen bestemmer altså observationernes værdier på den variable, som vi ønsker at estimere den kausale effekt af. At tærskelvariablen bestemmer interventionen deterministisk betyder, at en given intervention slås til eller fra ved en særlig tærskelværdi (*discontinuity* eller *cut-off*). Tærskelværdien er det punkt på tærskelvariablen, hvor observationerne overgår fra kontrol- til forsøgsgruppe. Den deterministiske opdeling betyder, at hvis vi kender observationernes værdier på tærskelvariablen, ved vi også med sikkerhed, om de hører til vores kontrol- eller forsøgsgruppe.

I tilfælde hvor tildelingen af forsøgs- eller kontrolstatus ikke er deterministisk, taler man om *fuzzy* RDD (Dunning, 2012). I fuzzy RDD kan observationer på samme side af tærskelværdien være havnet i både forsøgs- og kontrolgruppen. En observations placering omkring tærskelværdien giver derfor ikke længere sikker viden om observationens interventionsstatus. I disse tilfælde vil tærskelværdien alene fortælle os sandsynligheden for, at en observation tilhører enten kontrol- eller forsøgsgruppe (Angrist og Pischke, 2009: 259). Dvs. at tærskelværdien nu udgør et skarpt brud i *sandsynligheden* for at være tildelt en intervention. I disse tilfælde er der udviklet en række forskellige instrumentvariabelprocedurer til at estimere selektionsprocessen omkring tærskelværdien (se Hahn et al. (2001) og Klaauw (2002) for nærmere forklaring og eksempler). Her vil vi ikke behandle fuzzy RDD yderligere.

## Hvorfra kommer tærskelvariablene?

At finde egnede tærskelvariable er kernen i RDD. En tærskelvariabel kan lyde som et abstrakt fænomen, men er i virkeligheden en ret konkret og udbredt størrelse. Faktisk er der mange af politologiens centrale variable, som bestemmes af tærskelværdier. Som Green et al. (2009: 412) har argumenteret for, er der et stort potentiale for RDD i statskundskaben pga. "de rigide forhold hvorunder institutionelle regler fordeler repræsentation og regeringsressources". Med andre ord: I politisk-administrative sammenhænge er det ofte skarpe tærskelværdier, der bestemmer værdierne på nogle af de variable, som vi er aller mest interesserede i at estimere den kausale effekt af. Dunning (2012: 69)

inddeler mulige tærskelvariable i socialvidenskaberne inden for kategorierne: adgangseksamener, befolkningstærskler, størrelsesbaserede tærskler, tildelingskriterier, alderstærskler og tætte valg. Nogle af disse kategorier er imidlertid ikke klart adskilte. Et bud på nogle mere overordnede tærskelkategorier er derfor: output/outcome-information, geografi/rum, socioøkonomiske kriterier og tid/timing. Et kort kig på nogle eksempler fra litteraturen viser udmærket, hvor mange forskellige tærskelværdier der bestemmer politisk-administrative forhold inden for hver af disse kategorier:

Output/outcome-mål kan i en statskundskabssammenhæng dække over både valgregler og valgresultater, der deterministisk bestemmer, hvem der får politisk magt (Lee, 2008; Pettersson-Lidbom, 2008; Broockman, 2009; Olsen 2013b). Men det dækker også over resultatmålinger af offentlige ansatte eller organisationer, der bestemmer sanktioner og performance feedback i den offentlige sektor (Chiang, 2009; Hemelt, 2011; Olsen, 2013a). Lee (2008) var den første til at udnytte valgresultater som en tærskelvariabel. I mange valg-systemer ændrer udfaldet sig deterministisk, hvis en kandidat får over 50 pct. af stemmerne. I valg med to kandidater er der en verden til forskel på at få 49 pct. eller 51 pct. af stemmerne. Det er til gengæld ret ligegyldigt, om man får 48 pct. eller 49 pct. 50 pct. er altså den tærskelværdi, der deler sejrherrene fra de slagne.

Geografi dækker i statskundskaben typisk over observationernes afstand til grænser mellem nationer, kommuner eller andre former for administrative og politiske enheder (Black, 1999; Keele og Titunik 2013). Et godt eksempel er Chen et al. (2013), der benytter individers syd/nordplacering i forhold til Huai floden i Kina, der deterministisk giver gratis adgang til kulvarme og dermed også giver variation i graden af forurening, som borgerne oplever.

Socioøkonomiske kriterier kan være mål for befolkningsstørrelse, velstand eller sociale problemer, som bestemmer, hvem der har ret til at modtage hvilke ydelser (Ludwig og Miller, 2007; Elis et al., 2009; Becker et al., 2010; Hopkins, 2011). Ludwig og Miller (2007) har set på sundheds- og uddannelseseffekter af det amerikanske Head Start Program, der gav massiv støtte til børn i fattige familier. Det særlige ved programmet var, at hjælpen blev tildelt på baggrund af tærskler i amerikanske amters fattigdomsrater i 1960. Således var det kun familier i de 300 fattigste amter, der kunne modtage hjælp. Familier i det 301. fattigste amt var udelukket fra denne hjælp.

Endelig kan tærskelvariable udspringe af tidslig variation, der regulerer borgernes rettigheder i forhold til offentlige services såsom fødselsdato, alder, i særlige tidsintervaller eller lignende. For eksempel udnytter Almond og Doyle (2011), at lovgivningen i USA giver moderen ret til to døgn på hospitalet, hvor-

for børn født efter midnat i snit vil have et længere hospitalsophold end børn født umiddelbart før midnat.

## Fra tærskelvariable til estimater af kausale effekter

Hvordan kommer man fra tærskelvariablen og tærskelværdien til kausal identifikation? Den centrale antagelse er, at observationerne i nærheden af tærskelværdien ikke har haft mulighed for at selvselekttere sig ind på en særlig side af tærskelværdien, og at de heller ikke intentionelt er blevet placeret af andre. Placeringen af observationer omkring tærskelværdien skal således opfylde kriteriet om tilnærmelsesvis tilfældig (*as if random*) i nærheden af tærskelværdien (Dunning, 2012). Tænker vi for eksempel på Lee (2008), så er antagelsen, at i meget tætte valg er det tilfældige forhold, der afgør, hvem der vinder. I eksemplet med Almond og Doyle (2011) er det tanken, at det er tilfældigt, om man føder lige før eller efter midnat.

Hvis antagelsen om tilnærmelsesvis tilfældighed holder, betyder det, at observationer omkring tærskelværdien ikke er systematisk forskellige på observerbare og uobserverbare før-interventionsvariable. Dvs. der er ikke andre forhold givet før interventionen, som varierer systematisk omkring tærskelværdien, hvor interventionen slås til og fra. Det eneste, der varierer, er, om observationen tilhører kontrol- eller forsøgsgruppen, dvs. om de er placeret lige over eller under tærskelværdien.

At retfærdiggøre antagelsen om tilnærmelsesvis tilfældighed omkring tærskelværdien kræver både kvalitative argumenter og empiriske test. I hvilket som helst studie er det undersøgers opgave at redegøre for, hvorvidt RDD's antagelser holder i en given empirisk kontekst (Caughey og Sekhon, 2011: 405). Kvalitativt skal det være plausibelt, at selvselektion omkring tærskelværdien ikke er mulig. Det kan for eksempel begrundes med, at tærskelværdien har været ukendt for observationerne. Det vil gøre det mere plausibelt, at observationerne ikke har kunnet påvirke deres egen værdi på tærskelvariablen og dermed sikre, at de er kommet over eller under en særlig tærskelværdi. På samme måde kan det kvalitativt godtgøres, hvorvidt dem, der har bestemt tærskelværdien, har gjort det med henblik på, at særlige individer og organisationer skulle placeres på en bestemt side af tærskelværdien. Derfor er opgaven, som med alle andre naturlige eksperimentelle designs, at vi grundigt skal sætte os ind i den empiriske kontekst og de omstændigheder, der har tildelt observationerne værdier på tærskelvariablen (Dunning, 2012). Med Friedmans (1991) ord kan vi sige, at RDD er et godt eksempel på et design, der "slider på skosålerne": Det kræver, at vi nøje undersøger og forstår den empiriske kontekst.

Dette bør også ske ved kvantitativt, empirisk at undersøge, om der er forskelle mellem observationer i et snævert vindue omkring tærskelværdien. Dette kan ske ved at se på, om observationer varierer systematisk på før-interventionsvariable tæt omkring tærskelværdien. Hvis der ikke er systematiske forskelle på tværs af en række oplagte observerbare confounders, vil det styrke antagelsen om tilnærmelsesvis tilfældighed i placeringen af observationer i nærheden af tærskelværdien. Hvis data er tilgængelig for flere tidsperioder, bør man også teste, om observationerne er systematisk forskellige på den afhængige variabel og tærskelvariablen, i perioden før observationerne blev opdelt i kontrol- eller forsøgsgruppe (Caughy og Sekhon, 2011: 405). Hvis der er for eksempel er en systematisk forskel i den afhængige variabel for observationer omkring tærskelværdien, før denne værdi var kendt og trådte i kraft, kunne det tyde på, at netop denne værdi er valgt af strategiske grunde. Derudover bør man også undersøge, om der er forskydninger i fordelingen af observationer på tærskelværdivariablen omkring tærskelværdien (Imbens og Lemieux, 2008: 621). Hvis fordelingen er ujævn eller på andre måder uregelmæssig omkring tærskelværdien, kan det være et tegn på, at observationerne har haft mulighed for at påvirke deres placering omkring tærskelværdien. For eksempel hvis det af en eller anden grund er fordelagtigt at være i forsøgsgruppen, kunne man forestille sig, at observationer ville klumpe sig sammen lige på den rigtige side af tærskelvariablen, der giver adgang hertil. Dette kan både undersøges visuelt ved at plote frekvensen af observationer omkring tærskelværdien eller ved formelle test (McCray, 2008).

Der er oplagt, at der er mange tærskelværdier ude i verden, som deterministisk griber ind i folks liv, men hvor observationerne har gode muligheder for selvselektion eller at blive intentionelt placeret af andre. For eksempel er der massevis af tærskler i de fleste skattesystemer (eksempelvis topskattegrænser), men det er oplagt, at mennesker agerer strategisk i forhold til disse tærskler og har god mulighed for at påvirke deres egen indkomst i detaljen. Det er derfor ikke rimeligt at antage, at observationernes placering omkring tærskler af den type har været tilnærmelsesvis tilfældig. Det er altså ikke alle tærskler og tærskelvariable, der giver adgang til kausal inferens via RDD.

Hvis der er gode argumenter for at tro, at observationernes placering har været tilnærmelsesvis tilfældig, og hvis empiriske tests har kunnet underbygge dette, så kan vi sammenligne observationer omkring tærskelværdien for at få et estimat af den kausale effekt. Dvs. forskelle i gennemsnit på vores afhængige variabel omkring tærskelværdien udgør vores bedste bud på den kausale effekt af den intervention, der slås til ved tærskelværdien (Green et al., 2009). Spørgsmålet er så blot, hvordan vi kan estimere denne effekt.

## Statistisk modellering af RDD

Opgaven i det følgende er at give et introducerende overblik over, hvordan RDD-modeller estimeres i praksis. Detaljerede formelle præsentationer af RDD kan findes i Hahn et al. (2001), Imbens og Lemieux (2008) og Lee (2008). Udgangspunktet er en tærskelværdidummy, som er givet ved tærskelværdivariablen:

$$\text{tærskelværdidummy} = \begin{cases} 1, & \text{tærskelvariabel} > \text{tærskelværdi} \\ 0, & \text{tærskelvariabel} \leq \text{tærskelværdi} \end{cases}$$

I dette tilfælde er interventionen slået til, når tærskelvariablen har værdier, der er større end tærskelværdien, mens kontrolgruppen er defineret, når tærskelvariablen antager værdier, der er mindre end eller lig med tærskelværdien. Det er vigtigt, at tærskelvariablen er målt, før den afhængige variabel bestemmes, da tærskelvariablen skal bestemme fordelingen af den intervention, som vi ønsker at estimere den kausale effekt af. Vi kan nu opskrive RDD med afsæt i denne regressionsmodel:

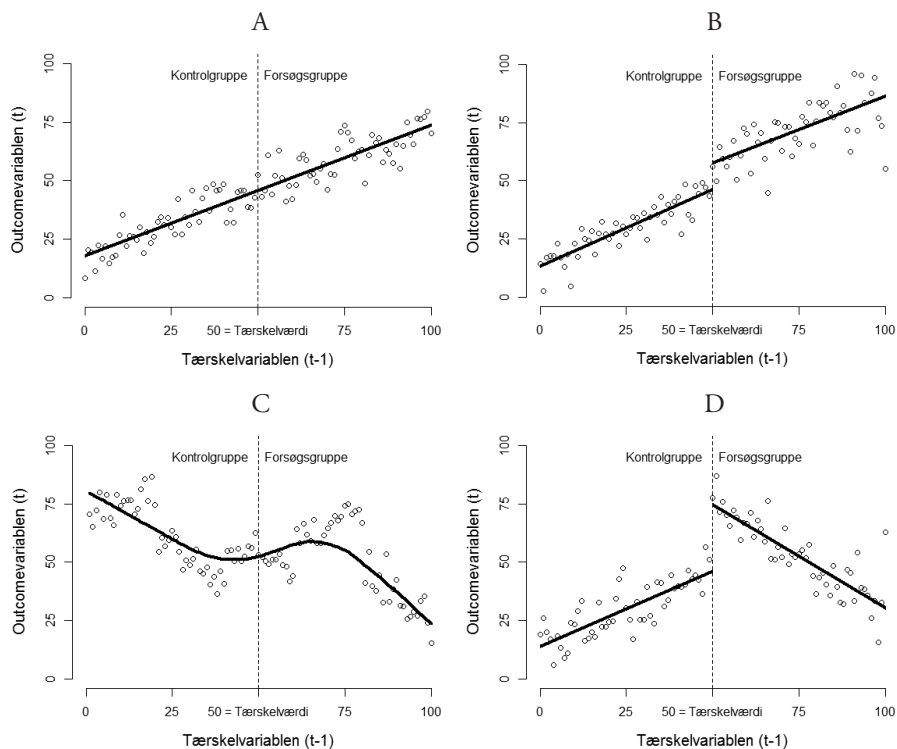
$$\text{outcome} = \alpha + \beta * \text{tærskelværdidummy} + f(\text{tærskelvariablen})$$

I modellen ovenfor er den afhængige variabel en funktion af tærskelværdidummen og tærskelvariablen.  $\beta$  for tærskelværdidummen angiver RDD-estimatet af den kausale effekt. Koefficienten angiver forskydningen i værdier på den afhængige variabel ved tærskelværdien. Det er denne "diskontinuitet", vi forsøger at estimere. Visuelt er det vist i figur 1 (med inspiration fra Mellor og Mark, 1998).

I panel A ser vi udgangspunktet for RDD. Vi har tærskelvariablen på x-aksen, som ved en given værdi tildeler observationer en kontrol- eller forsøgsstatus. I dette tilfælde er observationer med en tærskelværdi under 50 angivet som kontrolgruppe, mens observationer med en tærskelværdi med 50 eller over tilhører en forsøgsgruppe. Endelig har vi den afhængige variabel ud af y-aksen. I eksemplet forestiller vi os, at der er en positiv sammehæng mellem tærskelvariablen og den afhængige variabel. I panel A ser vi, at der ingen forskydning er i linjerne omkring tærskelvariablen, hvilket indikerer, at der ikke er nogen effekt. I panel B ser vi derimod et eksempel, hvor der er en klar forskydning efter tærskelværdien. I RDD er det i virkeligheden størrelsen på denne forskydning i linjerne, som man forsøger at estimere via  $\beta$  for tærskelværdidummen. Forskellen i forskydning i linjerne på tærskelværdien udgør den kausale effekt for observationer omkring tærskelværdien (Angrist og Pischke, 2009; Dunning, 2012). I det viste eksempel vil tærskelværdidummen være positiv svarende til



Figur 1:



en opadgående forskydning i linjen for observationerne i forsøgsgruppen over tærskelværdien.

En central udfordring i estimeringen af RDD-modeller ligger i at bestemme den funktionelle form på tærskelvariabelen og det vindue af data rundt om tærskelværdien, som indgår i analysen (Dunning, 2012). Problemet består i, at vi skal separere den forskydning, vi ønsker at estimere ved tærskelværdien fra den generelle sammenhæng, der ellers måtte være mellem tærskelvariabelen og den afhængige variabel. Det er ikke nogen let opgave. For det første har tærskelvariabelen ikke nødvendigvis en simpel lineær sammenhæng med den afhængige variabel. At specificere den korrekte funktionelle form på tærskelværdivariabelen er vigtigt for at undgå, at RDD-estimatet af den kausale effekt ikke blot afspejler ikke-linearitet omkring tærskelværdien (Angrist og Pischke, 2009: 254). En måde at gøre dette på er ved at tilføje forskellige sæt af polynomier af tærskelværdivariabelen (Green et al., 2009: 405). Derved tillades forskellige ikke-lineære sammenhænge mellem tærskelværdivariabelen og den afhængige

variabel. I panel C er vist et eksempel, hvor der er en ikke-lineær sammenhæng mellem tærskelvariablen og den afhængige variabel. I dette tilfælde er det tydeligt, at hvis den funktionelle form ikke blev modelleret tilstrækkeligt fleksibelt, ville man foranlediges til at tro, at der var en kausal effekt omkring tærskelværdien. Derudover bør man tillade den funktionelle form at variere på hver side af tærskelværdien. I panel D er vist et eksempel, hvor tærskelvariablens sammenhæng med den afhængige variabel varierer rundt om tærskelværdien. Det er et udtryk for, at sammenhængen mellem tærskelvariablen og den afhængige variabel ændres, idet interventionen slås til. Derfor kan det være nødvendigt at modellere en interaktion mellem tærskelvariablen og tærskelværdien.

Ud over at variere den funktionelle form for tærskelvariablen skal man også bestemme det vindue af data omkring tærskelværdien, som medtages i analysen. Dvs. man skal variere, hvor stor afstand man vil tillade, at observationer har til tærskelværdien. Herved gør man resultatet mindre overfølsomt over for, om den rigtige funktionelle form er specificeret for tærskelvariablen (Dunning, 2012). Dvs. i takt med at vi indsnævrer vinduet af data omkring tærskelværdien, bør vi kunne modellere tærskelværdivariblen stadig mere enkelt. Der er oplagt, at jo mere vi indsnævrer det vindue af data, som der analyseres på, omkring tærskelværdien, desto mere immune bliver vi over for mulige confounders og fejlspecificeret funktionelform for tærskelvariablen (Green et al., 2009: 401). Det er kun helt tæt omkring tærskelværdien, at idéen om tilnærmelsesvis tilfældighed er plausibel. Det er nemlig mere sandsynligt, at observationer, der ligger meget langt over eller under tærskelværdien, er systematisk forskellige fra hinanden (Dunning, 2012: 127). Ved at udvide vinduet kan vi derfor introducere en bias i vores estimat af den kausale effekt, som stammer fra faktorer, der korrelerer med vores intervention. Omvendt er det ikke uden problemer at begæne vinduet af data, som der analyseres på. Ved at gøre vinduet smallere mister vi en masse data, hvilket øger usikkerheden på den estimerede effekt (Green et al., 2009: 412). En mulighed er at bruge mere avancerede ikke-parametriske estimationsmetoder, som er blevet udviklet til statistikprogrammer som R og STATA. Her benyttes oftest såkaldt lokal linear regression og procedurer til at bestemme det optimale vindue af data omkring tærskelværdien, som analyserne skal baseres på (Imbens og Lemieux, 2008; Lee og Lemieux, 2009; Imbens og Kalyanaraman, 2009). Dunning (2012: 132) har dog omvendt argumenteret for, at man altid som udgangspunkt bør estimere simple forskelle i gennemsnit for et sæt af observationer, der er placeret så tæt på tærskelværdien som muligt. Argumentet er, at hvis der er rigeligt med data omkring tærskelværdien, og hvis antagelsen om ”tilnærmelsesvis tilfældig” er velbegrunder, så har vi at gøre med et lokalt eksperiment for observationer

tæt ved tærskelværdien. Af samme grund kan den kausale effekt estimeres ved at estimere en simpel forskel i gennemsnit for observationer over og under tærskelværdien. Kun i tilfælde hvor data er sparsomme tæt ved tærskelværdien, bliver vi nødsagede til at modellere den funktionelle form eller vægte observationerne efter deres nærhed til tærskelværdien for at kunne inkludere data længere væk (Dunning, 2012: 133-134). I alle tilfælde er det en god idé at estimere forskydning i linjerne omkring tærskelværdien ved så mange forskellige metoder som muligt.

## Eksempler på brug af RDD i valgforskning og forvaltning

I det næste skal vi se mere specifikt på to konkrete eksempler på studier, der anvender RDD. Det ene kommer fra valgforskningen og omhandler *incumbency advantage* i den amerikanske kongres (Lee, 2008). Der er langt fra det eneste eksempel på RDD-studier af de vælgermæssige fordel ved at besidde magten (Leigh, 2008; Butler, 2009; Caughey og Sekhon, 2011; Gerber og Hopkins, 2011). Det andet eksempel stammer fra forskningen i offentlig forvaltning og politik og omhandler effekter af negative resultatmålinger på skolers performance (Chiang, 2009). Dette er blot ét blandt flere studier, der benytter RDD til at studere, effekten af resultatinformation på offentlige organisationers performance (Hemelt, 2011; Olsen, 2012, 2013a).

### *Eksempel 1: Incumbency advantage i den amerikanske kongres (Lee, 2008)*

I amerikansk forskning har der været en stor interesse for at estimere den såkaldte *incumbency advantage*: dvs. er der en vælgermæssig fordel af at besidde magten? Rent deskriptivt er succesraten for genvalg i den amerikanske kongres på omkring 90 pct. Dette er dog næppe et udtryk for en ren *incumbency effect*: Politikere, der har opnået valg, er systematisk forskellige fra de kandidater, som de i sin tid vandt over. De samme kvaliteter, som gjorde det muligt for dem at vinde pladsen i første omgang, vil også påvirke deres genvalgschancer og kvaliteten af potentielle udfordrere i fremtidige valg. Nogle af disse forskelle kan måske observeres, og der kan derfor tages højde for dem i analysen, mens mange andre forhold (eksempelvis kandidat-kvalitet) ikke umiddelbart kan observeres. Som tidligere nævnt foreslår Lee (2008), at tætte valg opfylder betingelserne i RDD. Ved at sammenligne vindere og tabere i tætte valg opnår vi en kontrol- og forsøgsgruppe for *incumbency advantage*, hvor der ikke er systematiske forskelle på hverken observerbare eller uobserverbare forhold. Ved at sammenligne valgresultater i senere valg mellem vindere og tabere opnår vi således et estimat af den kausale effekt af *incumbency advantage*. Lee (2008) finder på den baggrund, at der er en positiv *incumbency advantage*, som øger

stemmeandelen i senere valg med omkring 7-8 procentpoint. I et nyere studie har Caughey og Sekhon (2011) dog vist, at studier med RDD af tætte valg til den amerikanske kongres, herunder Lee (2008), formentlig ikke opfylder kravet om fravær af selvselektion for observationer tæt ved tærskelværdien. Således er vindere og tabere systematisk forskellige fra hinanden i selv helt tætte valg. Vindere har større politisk erfaring, flere økonomiske ressourcer, og politiske iagttagere har før valget vurderet de senere vindere som værende bedre (Caughey og Sekhon, 2011: 404). Eksemplet viser, at selv i situationer, hvor antagelserne bag RDD virker plausible, så kan nærmere undersøgelser vise, at de i realiteten ikke holder. Men det viser også, at RDD's antagelser er så relativt simple at teste, at vi hurtigt kan identificere problemer for kausal inferens via RDD. I tillæg hertil viser Eggers et al. (2013), at USA måske er en anomali, da antagelserne for RDD er opfyldt i analyser af 40.000 valg fra syv forskellige demokratier.

### *Eksempel 2: Effekter af negative resultatmålinger på skolers performance (Chiang, 2009)*

Inden for forvaltningsvidenskaben har meget fokus været rettet mod, hvordan resultatmålinger påvirker borgere, offentlige organisationer og de offentlige ansatte. Problemet er blot, at spørgsmålet om den kausale effekt af forskellige typer performance feedback er meget kompliceret. Organisationer klarer sig på forskellig vis af gode grunde, og de selv samme grunde vil formentlig også spille ind på, hvordan organisationerne reagerer på at blive målt og vejet. Man kan for eksempel forestille sig, at en skole får at vide, at den er blandt landets dårligste målt på karakterer. De selv samme faktorer, som i første omgang fik skolen til at klare sig dårligt, kan også tænkes at påvirke, hvordan skolen reagerer på at få at vide, at den er dårlig. Vi har hverken en teoretisk model eller de nødvendige data til at tage højde for de potentielle confounders, der kan være, når vi ønsker at slutte fra variation i resultatmålinger og til efterfølgende organisatorisk adfærd (Olsen, 2012, 2013a).

Chiang (2009) foreslår RDD som en løsning til at estimere kausale effekter af negative resultatmålinger på offentlige organisationers adfærd i en amerikansk kontekst. I staten Florida tildeles skolerne en overordnet bogstavskarakter på baggrund af tærskelværdier i et samlet kontinuert mål for elevernes resultater. Hvis skolerne på denne tærskelvariabel opnår mindre end 280 point, så tildeles de et skamfuldt "F", som er ranglistens dårligste kategori. Hvis en skole får over 280 point tildeles den karakteren "D". Ved at sammenligne skoler lige omkring tærskelværdien på 280 point, opnår Chiang eksogen variation i de overordnede resultatmål, som den enkelte skole har opnået. Hvis en skole

får karakteren “F” får den blandt andet trusler om en lang række sanktioner, hvis ikke resultaternes forbedres. F-skoler har derfor stærk tilsyndelse til at forbedre sig. Det er ikke umiddelbar plausibelt, at skolerne kan selvselekttere sig ind på en mere fordelagtig karakter i nærheden af tærsklen. Skolerne omkring tærskelværdien er derfor af samme grund ens på en lang række observerbare forhold. Således kan Chiang (2009) vise, at truslen om sanktioner sætter sig igennem via bedre resultater i matematik til børnene i F-skoler. Han viser ligeledes, at F-skoler i højere grad end D-skoler investerer i undervisningsteknologi og efteruddannelse af lærerpersonalet. Studiet er et godt eksempel på, hvordan RDD kan hjælpe med at besvare et ellers kompliceret kausalt spørgsmål i studiet af offentlig forvaltning og politik.

## Begrænsning og udfordringer for RDD i statskundskab – og nogle løsninger

Afslutningsvis er det værd at overveje, hvilke barrierer og udfordringer der kan være for, at statskundskaben kan benytte RDD.

For det første er det oplagt, at politiske og administrative aktører med de rette incitamenter vil gøre alt for at placere sig strategisk omkring vigtige tærskelværdier. Dette er et problem i forhold til den helt centrale antagelse om, at observationerne ikke har mulighed for at selvselekttere sig “til rette” omkring tærskelværdier (Cook, 2008). For politologer er der derfor en særlig pligt til både kvalitativt og kvantitativt, empirisk at vise, at tærskelværdier ikke bliver strategisk manipuleret af de undersøgte observationer eller af andre med interesser i placeringen omkring tærskelværdien.

For det andet kan der i politisk-administrative sammenhænge være problemer med at opnå eksakt viden om tærskelværdier og tærskelvariable. Typisk vil disse informationer ikke være offentligt tilgængelige, og det vil kræve tålmodighed at grave dem frem, hvis det overhovedet er muligt. Nogle gange kan tærskelværdier have mere uformel karakter, eller aktører kan have en interesse i at holde dem for sig selv. Et tilgrænsende problem kan være, at der i politisk-administrative sammenhænge kan være tale om uklare og derfor ikke-deterministiske tærskelværdier, hvilket betyder, at man må ty til fuzzy RDD. For politologer vil det derfor ofte kræve indgående empiriske viden og tålmodighed at identificere de nødvendige informationer, for at RDD kan implementeres korrekt.

For det tredje er RDD kun i stand til at estimere en lokal kausal effekt, der gælder for observationer omkring tærskelværdien (Imbens og Lemieux, 2008: 621). Dvs. at RDD's høje interne validitet og få antagelser har den omkostning, at estimatet af den kausale effekt er begrænset til den subpopulation af data,

som ligger omkring tærskelværdien. Udfordringen for politologer er derfor at beskrive, i hvilket omfang estimatet af den kausale effekt kan forventes at se ud for observationer længere væk fra tærskelværdien. Hvad kan vi forvente, at der vil ske med den kausale effekt i takt med, at vi bevæger os væk fra tærskelværdien? Man bør altid overveje om særlige forhold gør sig gældende for observationer omkring tærskelværdien. Dette er dog ikke væsensforskelligt fra de fleste eksperimentelle designs, hvor estimater altid vil være lokale i forhold til en særlig population, kontekst eller tid (Linden et al., 2006; Angrist og Pischke, 2009).

For det fjerde er RDD relativt inefficent og kræver derfor store mængder data for at kunne identificere signifikante effekter. Dette skyldes, at modellerne oftest kun estimeres på data med en given afstand til tærskelværdien, og at tærskelværdiindikatoren og tærskelværiablen vil være stærkt korrelerede, hvilket vil forstørre standardfejlene (Green et al., 2009). Flere har blandt andet vist, at et typisk RDD-studie kræver tre til fire gange så meget data for at vise samme effektstørrelse som i et tilsvarende traditionelt eksperiment (Goldberger, 1972; Schochet, 2009). Politologer må derfor tænke i alternative dataformer som pannedata eller multi-level data, hvilket kan give adgang til flere observationer (Chiang, 2009; Pennell et al., 2011)

## Konklusion

Denne artikel har givet en introduktion till RDD for statskundskabere. Med RDD har vi et stærkt design, som med få og testbare antagelser gør det muligt at besvare kausale spørgsmål uden aktiv intervention. På mange måder er RDD lige så intuitivt og simpelt som det traditionelle eksperiment, men vi undgår de etiske og politiske problemer, som forkserintervention oftest kan medføre i statskundskaben (Campbell, 1969). Med RDD er sammenligningen af observationer omkring tærskelværdier sat i stedet for aktiv forskerintervention via lodtrækning som middel til opdeling af observationer i kontrol- og forsøgsgrupper. Sammenligninger med tilsvarende kontrollerede eksperimenter har vist, at RDD har en forbløffende evne til at fremkomme med sammenlignelige estimater (der dog er en del mindre effiente). Endelig er RDD et af de meget få designs til kausale analyser, som udspringer af socialvidenskaberne (Thistlethwaite og Campbell, 1960; Cook, 2008).

Der er derfor god grund til at se nærmere på RDD for statskundskaben. Den virkelige udfordring ligger i at finde egnede empiriske situationer. Det er i virkeligheden her, at forskerens primære arbejdsopgave forskyder sig hen i brugen af RDD. Er der tærskelvariable og tærskelværdier, som inddeler de observationer, man ønsker at studere, i kontrol- og eksperimentgrupper? Er det

kvalitativt plausibelt, at observationernes placering i nærheden af tærskelværdien har været tilnærmelsesvis tilfældig? Og kan man vise, at observationerne i nærheden af tærskelværdien faktisk ikke er systematisk forskellige på observerbare confounders? Der er de spørgsmål, som alle, der ønsker at anvende RDD til kausale spørgsmål, først og fremmest må beskæftige sig med.

## Note

1. Stor tak til redaktørerne Mogens Kamp Justesen, Robert Klemmensen og Kim Mannemar Sønderskov samt to anonyme læsere for at komme med meget konstruktive kommentarer til manus. Også tak til Ulrik Hvidman for nogle gode ændringsforslag. Eventuelt tilbageværende uklarheder og fejl er mine egne alene.

## Litteratur

- Aiken, Leona S., Stephen G. West, David E. Schwalm, James L. Carroll og Shenghwa Hsiung (1998). Comparison of a randomized and two quasi-experimental designs in a single outcome evaluation. *Evaluation Review* 22 (2): 207-244.
- Almond, Douglas og Joseph J. Doyle (2011). After midnight: A regression discontinuity design in length of postpartum hospital stays. *American Economic Journal: Economic Policy* 3 (3): 1-34.
- Angrist, Joshua D. og Jörn-Steffen Pischke (2009). *Mostly Harmless Econometrics*. Princeton, MA: Princeton University Press.
- Becker, Sascha O., Peter H. Egger og Maximilian von Ehrlich (2010). Going NUTS: The effect of EU Structural Funds on regional performance. *Journal of Public Economics* 94 (9-10): 578-590.
- Berk, Richard, Geoffrey Barnes, Lindsay Ahlman og Ellen Kurtz (2010). When second best is good enough: a comparison between a true experiment and a regression discontinuity quasi-experiment. *Journal of Experimental Criminology* 6 (2): 191-208.
- Black, Sandra E. (1999). Do Better Schools Matter? Parental Valuation of Elementary Education. *The Quarterly Journal of Economics* 114 (2): 577-599.
- Broockman, David E. (2009). Do Congressional Candidates Have Reverse Coattails? Evidence from a Regression Discontinuity Design. *Political Analysis* 17 (4): 418-434.
- Butler, Daniel M. (2009). A regression discontinuity design analysis of the incumbency advantage and tenure in the U.S. House. *Electoral Studies* 28 (1): 123-8.
- Campbell, Donald T. (1969). Reforms as experiments. *American Psychologist* 24 (4): 409-429.
- Caughey, Devin og Jasjeet S. Sekhon (2011). Elections and the regression discontinuity design: Lessons from close US house races, 1942-2008. *Political Analysis* 19(4): 385-408.

- Chen, Yuyu, Avraham Ebenstein, Michael Greenstone og Li Hongbin (2013). Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy. *Proceeding of the National Academy of Science* 110 (32): 12936-12941.
- Chiang, Hanley (2009). How accountability pressure on failing schools affects student achievement. *Journal of Public Economics* 93 (9-10): 1045-1057.
- Cook, Thomas (2008). Waiting for Life to Arrive? A history of the regression-discontinuity design in Psychology, Statistics and Economics. *Journal of Econometrics* 142 (2): 636-654.
- Dunning, Thad (2012). *Natural Experiments in the Social Sciences. A Design-Based Approach*. Cambridge, UK: Cambridge University Press.
- Eggers, Andrew, Ole Folke, Anthony Fowler, Jens Hainmueller; Andrew B. Hall, James M. Snyder (2013). On the validity of the regression discontinuity design for estimating electoral effects: New evidence from over 40,000 close races. *Working paper*.
- Elis, Roy, Neil Malhotra og Marc Meredith (2009). Apportionment cycles as natural experiments. *Political Analysis* 17 (4): 358-376.
- Freedman, David A. (1991). Statistical models and shoe leather. *Sociological Methodology* 21: 291-313.
- Gerber, Elisabeth R. og Daniel J. Hopkins (2011). When mayors matter: Estimating the impact of mayoral partisanship on city policy. *American Journal of Political Science* 55 (2): 326-339.
- Goldberger, Arthur S. (1972). Selection bias in evaluating treatment effects: The case of interaction. Technical report, Institute for Research on Poverty, University of Wisconsin, Madison.
- Green, Donald P., Terence Y. Leong, Holger L. Kern, Alan S. Gerber og Christopher W. Larimer (2009). Testing the Accuracy of Regression Discontinuity Analysis Using Experimental Benchmarks. *Political Analysis* 17 (4): 400-417.
- Hahn, Jinyong, Petra Todd og Wilbert Van der Klaauw. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica* 69 (1): 201-209.
- Hemelt, Steven W. (2011). Performance effects of failure to make Adequate Yearly Progress (AYP): Evidence from a regression discontinuity framework. *Economics of Education Review* 30 (4): 702-723.
- Hopkins, Daniel J. (2011). Translating into votes: The electoral impacts of Spanish-language ballots. *American Journal of Political Science* 55 (4): 814-830.
- Imbens, Guido og Karthik Kalyanaraman (2009). Optimal bandwidth choice for the regression discontinuity estimator. Unpublished manuscript, Department of Economics, Harvard University.



- Imbens, Guido og Thomas Lemieux (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics* 142 (2): 615-635.
- Keele, Luke og Rocio Titiunik (2013). Natural experiments based on geography. *Working paper*.
- Klaauw, Wilbert van der (2002). Estimating the effect of financial aid offers on college enrollment: A regression-discontinuity approach. *International Economic Review* 43 (4): 1249-1287
- Lee, David S. (2008). Randomized experiments from non-random selection in U.S. House elections. *Journal of Econometrics* 142 (2): 675-697.
- Lee, David S. og Thomas Lemieux (2009). Regression discontinuity designs in economics. *National Bureau of Economic Research Working Paper*.
- Leigh, Andrew (2008). Estimating the impact of gubernatorial partisanship on policy settings and economic outcomes: A regression discontinuity approach. *European Journal of Political Economy* 24 (1): 256-268.
- Linden, Ariel, John L. Adams og Nancy Roberts (2006). Evaluating disease management programme effectiveness: an introduction to the regression discontinuity design. *Journal of Evaluation in Clinical Practice* 12 (2): 124-131.
- Ludwig, Jens og Douglas L. Miller (2007). Does head start improve children's life chances? Evidence from a regression discontinuity design. *Quarterly Journal of Economics* 122 (1): 159-208.
- McCray, Justin (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics* 142 (2): 698-714.
- Mellor, Steven og Melvin M. Mark (1998). A quasi-experimental design for studies on the impact of administrative decisions: Applications and extensions of the regression-discontinuity design. *Organizational Research Methods* 1 (3): 315-333.
- Morgan, Stephen L. og Christopher Winship (2007). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge: Cambridge University Press.
- Olsen, Asmus L. (2012). Regression discontinuity designs in public administration: The case of performance measurement research. Paper, ECPR Joint Sessions, Antwerpen, 11.-15. april.
- Olsen, Asmus L. (2013a). Naming bad performance: Can performance disclosure drive improvements? Revise and resubmit. *Journal of Public Administration Research and Theory*.
- Olsen, Asmus L. (2013b). Incumbency advantage and tax and spending effects of a unified leadership: A regression discontinuity design among Danish municipalities. Ufærdigt manus.

- Papay, John P., John B. Willett og Ricard J. Murnane (2011). Extending the regression-discontinuity approach to multiple assignment variables. *Journal of Econometrics* 161 (2): 203-207.
- Pennell, Micheal, Erinn M. Hade, David M. Murray og Dale A. Rhoda (2011). Cutoff designs for community-based intervention studies. *Statistics in medicine* 30 (15): 1865-1882.
- Pettersson-Lidbom, Per (2008). Do parties matter for economic outcomes? A regression-discontinuity approach. *Journal of the European Economic Association* 6 (5): 1037-1056.
- Robinson, Gregory, John E. McNulty og Jonathan S. Krasno (2009). Observing the counterfactual? The search for political experiments in nature. *Political Analysis* 17 (4): 341-357.
- Rubin, Donald B. (1977). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Statistics* 2 (1): 1-26.
- Schochet, Peter Z. (2009). Statistical power for regression discontinuity designs in education evaluations. *Journal of Educational and Behavioral Statistics* 34 (2): 238-266.
- Sekhon, Jasjeet S. (2010). Opiates for the matches: Matching methods for causal inference. *Annual Review of Political Science* 12: 487-508.
- Shadish, William R. (2011). The truth about validity. *New Directions for Evaluation* (130): 107-117.
- Shadish, William R., Thomas D. Cook og Donald T. Campbell (2001). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Stamford, Connecticut: Cengage Learning (2. udgave).
- Thistlethwaite, Donald L. og Donald T. Campbell. (1960). Regression-discontinuity analysis: An alternative to the ex post factor experiment. *The Journal of Educational Psychology* 51 (6): 309-317.