



Optagelseskvotens betydning for frafaldssandsynligheden

En empirisk analyse af universitetsbachelorstuderende
i perioden 2016-2018

KANDIDATAFHANDLING

Sabine Varning Lykkegaard Hansen

Dato: 31. August, 2020

Antal tegn: 128.921

ETCS-point: 30

Vejleder: Mette Gørtz



Abstract

This paper examines the effect of admission quotas on probability of dropout of Danish university students. The admission process into higher education in Denmark creates credible instruments from discontinuities that randomize applicants near an unpredictable GPA-based admission cutoff into different locations. Using a regression discontinuity (RD) design this paper estimates the causal effect of the admission quotas on the early dropout for students at Danish universities. Using Danish administrative data on Danish university students from 2016 to 2018, I show that the quotas does have an effect on probability of dropout and admission through quota 2 minimize dropout probability compared to admission through quota 1. This result could imply that the Danish admission system gives the students more motivation to finish their education after getting an admission through quota 2. Furthermore, the paper does investigate the heterogeneous effects of the results. The results implies that it is effective to increase the admission through to secure a lower dropout rates.

Tilkendegivelser

Jeg vil gerne takke Danmark Evalueringsinstitut, EVA, for samarbejdet om mit speciale. Jeg er dybt taknemmlige for, at Danmark Evalueringsinstitut har stillet data til rådighed, så det har været muligt at foretage analyserne. Derudover skal de have tak den altid gode rådgivning og sparring undervejs i processen og for at have stillet en kontorplads til rådighed. Endelig skal Mette Gørtz have et stort tak, for en meget værdsat og god vejledning.

Indhold

Resumé	iv
Figurer	vi
Tabeller	vi
1 Indledning	1
2 Litteratur	4
2.1 Internationale studier	4
2.2 Danske studier	4
3 Teori	7
3.1 Hvorfor tage en videregående uddannelse?	7
3.2 Hvorfor vælger nogle studerende at afbryde deres studie?	8
3.3 Teoretisk hypotese	9
4 Det danske optagesystem	11
4.1 Den Koordinerede Tilmelding	11
4.2 Den sociale mobilitet i optagesystemet	13
5 Data	15
5.1 Data kilder og variable	15
5.2 Data restriktioner og population	21
5.3 Deskriptiv statistik	23
6 Empirisk strategi	26
6.1 Regression diskontinuitet design	26
6.2 Økonometrisk specifikation	33
6.3 Udeladte variable	34
7 Resultater	37
7.1 Intern validitet af det empiriske design	37

7.2	Optagelseskvotens betydning på frafaldssandsynligheden	39
7.3	Heterogenitet i treatment-effekter	43
8	Robusthedscheck	49
8.1	Fordelingen af afstandsvariablen	49
8.2	Båndbreddens betydning	52
8.3	Forudbestemte kovariater	53
9	Diskussion	55
9.1	Begrænsninger	55
9.2	Politiske interventioner	57
9.3	Mulige fremtidige undersøgelser	59
10	Konklusion	61
11	Litteraturliste	62

Figurer

6.1	Betinget sandsynlighed for at modtage behandling i det skarpe RD design	29
6.2	RD behandlingseffekt i det skarpe RD design	30
7.1	Sammenhængen mellem optagelseskvote og frafaldssandsynligheden for universitetsbachelorer, 2016-2018	38
7.2	Optagelseskvotes effekt på førsteårs frafaldssandsynligheden for universitetsbachelorer, 2016-2018	40
7.3	Optagelseskvotes effekt på førsteårs frafaldssandsynligheden for universitetsbachelorer, 2016-2018	41
8.1	Fordeling af afstandvariabel, 2016-2018	50
8.2	Sensitiviteten af estimater i forhold til valget af båndbredden	52
.1	Optagelseskvotes effekt på førsteårs frafaldssandsynligheden for universitetsbachelorer med passiv uden motiveret ansøgning, 2016-2018	78
.2	Optagelseskvotes effekt på førsteårs frafaldssandsynligheden for universitetsbachelorer med kvote 2 med passiv med motiveret ansøgning, 2016-2018	79
.3	Optagelseskvotes effekt på førsteårs frafaldssandsynligheden for universitetsbachelorer med kvote 2 med fremmøde, 2016-2018	80

Tabeller

5.1	Data kilder	15
5.2	Variabeloversigt	21
5.3	Data restriktioner	22
5.4	Deskriptiv statistik	24
7.1	Effekten af optagelseskvoten på førsteårs frafaldssandsynlighed for universitetsbachelorer, 2016-2018	42
7.2	Heterogene effekter af optagelseskvoten på studerendes frafaldssandsynlighed for subpopulationer defineret ved baggrundskaraktistika	44
7.3	Heterogene effekter af optagelseskvoten på studerendes frafaldssandsynlighed for subpopulationer defineret ved uddannelseskaraktistika	46
7.4	Heterogene effekter af optagelseskvoten på studerendes frafaldssandsynlighed for subpopulationer defineret ved uddannelseskaraktistika	47

8.1	Test af kovariat	53
.1	Balancetabel	74
.2	Effekten af optagelseskvoten på førsteårs frafaldssandsynlighed for universitetsbachelorer for det simple RD design med forskellige båndbredde, 2016-2018	76
.3	Effekten af optagelseskvoten på førsteårs frafaldssandsynlighed for universitetsbachelorer for RD designet med fixed effects for forskellige båndbredde, 2016-2018	76
.4	Effekten af optagelseskvoten på førsteårs frafaldssandsynlighed for universitetsbachelorer for RD designet med fixed effects og demografiske kontroller for forskellige båndbredde, 2016-2018	77
.5	Effekten af optagelseskvoten på førsteårs frafaldssandsynlighed for universitetsbachelorer for RD designet med fixed effects, demografiske kontroller og selvrapporteret afklarethed for forskellige båndbredde, 2016-2018	77

1 Indledning

Ifølge OECD forlader omtrent en fjerdedel af de studerende på en videregående uddannelse deres studie før de får deres bachelor. (OECD, 2020) Disse tal dækker over store variationer fra lande med gennemførelsesprocenter på mindre end 60% (f.eks. Chile og Tyrkiet) til andre på over 90% (f.eks. Belgien og Finland). Rapporten viser ligeledes at uddannelsessystemet producerede næsten dobbelt så mange kandidater i 2007 som i midten af 1990'erne og dermed har ekspanderet kraftigt. I en dansk kontekst, har der i en årerække været en dansk politisk målsætning om, at få flere unge igennem en uddannelse, som særlig har været fordret af den øgede globalisering. Den politiske målsætning har været, at 60% af en ungdomsårgang skulle i 2020 tage en videregående uddannelse, og det øgede politiske fokus har også afspejlet sig i antallet af ansøgere til de videregående uddannelser. Tilgangen af optagne til de videregående uddannelser er de seneste 10 år steget med 32% med 52.552 studerende i 2010 til 69.526 studerende i 2020. (Uddannelses- og Forskningsministeriet, 2020)

Med det stigende antal studerende har det været et øget fokus på indretningen af uddannelsessystemet og selve optaget til de videregående, herunder hvordan kvaliteten og relevansen af uddannelserne sikres. Derudover har der været et fokus på, hvem der optages på de videregående uddannelser, og om optagesystemet er retfærdigt. Samt hvordan man sikre, at de studerende får den mest relevante uddannelse hurtigt mulig med øje for det fremtidige arbejdsmarkedet, hvilket er væsentligt, da flere studerende på videregående uddannelse alt andet lige betyder øgede offentlige udgifter via blandt andet taxametersystemet og Statens Uddannelsesstøtte (SU). Hertil kommer, at når unge begynder en videregående uddannelse, skifter de fra at være i fra beskæftigelse eller på offentlig forsørgelse. Endelig er forventningen, at de unge efter endt uddannelse kan indtræde på arbejdsmarkedet til en væsentlig højere indkomst sammenlignet med før uddannelsen, hvorfor uddannelse er en god investering for samfundet. (Uddannelses- og Forskningsministeriet, 2017) Der er således en samfundsøkonomisk interesse i tiden det tager de unge at få en uddannelse og hvilken uddannelse de får. Matchet mellem uddannelse og den enkelte studerende kan tænkes at have en effekt gennem flere forskellige kanaler. Eksempelvis kan matchet have en betydning for den viden og de evner den studerende opbygger i studietiden, tidspunktet for studiestart samt tiden det tager den studerende at gennemføre

studiet.

Der er gratis adgang til de danske videregående uddannelser forudsat at ansøgeren har en relevant ungdomsuddannelse. Der er således ingen økonomiske restriktioner i form af studieafgifter eller lignende. I Danmark foregår matchningen ved hjælp af et optagelsessystemet, der benyttes til fordelingen af studiepladser. Dette skyldes, at der er kapacitetsbegrænsninger på en lang række af uddannelser, som betyder, at ikke alle ansøgere bliver optaget på den uddannelse, som de helst vil ind på. Pladserne på uddannelser med kapacitetsbegrænsninger vil for mange uddannelser blive fordelt i to kvoter. En kvote, kvote 1, hvor pladserne vil blive fordelt efter adgangskvotienter, hvor ansøgere med de højeste adgangskvotienter vil blive tilbudt en studieplads, og en anden kvote, kvote 2, hvor de resterende pladser vil blive fordelt efter mere subjektive kriterier, som f.eks. motiveret ansøgning eller optagelsesprøve.

Jeg ønsker med dette speciale at undersøge, hvilke betydning at optagelseskvoten har for den studerendes vedholdenhed. Nærmere bestemt ønsker jeg at estimere, hvilken effekt optagelseskvoten har for universitetsbachelorstuderendes studiefrafald, som sker i løbet af det første studieår. På baggrund af overstående vil jeg i dette speciale undersøge følgende forskningsspørgsmål:

"Hvilken effekt har optagelseskvoten på sandsynligheden for tidligt frafald blandt universitetsbachelorstuderende?"

For at besvare dette forskningsspørgsmål udnyttes de institutionelle rammer for det danske optagelsessystem til universitetsuddannelserne for at undersøge den kausale sammenhæng mellem optagelseskvoten og sandsynligheden for tidligt frafald. Her anvendes regler for optagelse på de videregående uddannelser skaber en eksogen variation i optagelsen af ansøgere. Reglerne for optagelse på lange videregående uddannelser skaber en eksogen variation i optagelse, herunder optagelse i kvote 1.

Ideen er at bruge ansøgere til uddannelser, hvor optagelsen er begrænset af en adgangskvotient, hvilket vil sige, at der for den pågældende uddannelse, er der flere ansøgere, end der er pladser. På ansøgningstidspunktet ved ansøgerne ikke, hvad adgangskvotienten vil blive det år, hvorfor de ikke på forhånd kan forudsige, om de bliver optaget eller ej. Hvis adgangskvotienten viser sig at være under eller lig en ansøgers karaktergennemsnit, vil sandsynligheden for at blive optaget være større end hvis karaktergennemsnittet er under adgangskvotienten. Denne usikkerhed

medfører, at nogle ansøger både søge optagelse i kvote 1 og kvote 2 for at være bedre stillet. Dermed udnyttes det, at ansøgerne ikke har perfekt kontrol over, hvilken kvote som de optages i.

Som følge af resultaterne på ovenstående forskningsspørgsmål vil jeg diskutere om, skal fordelingen af studiepladser skal øges for den ene kvote fremfor den anden, og om det danske kvotesystem på videregående uddannelser fungerer efter hensigten, hvor studiepladserne bliver fordelt optimalt og retfærdigt? Eller om det er nødvendigt, at foretages en politisk indgriben, således at der sikres at færre studerende falder fra.

Specialet er struktureret på følgende vis: Afsnit 2 giver et litteraturoverblik, afsnit 3 giver en kort introduktion til de teoretiske rammer for specialet, afsnit 4 beskriver de institutionelle rammer for det danske optagelsesystem, afsnit 5 gennemgår de data der anvendes for at kunne besvare mit forskningsspørgsmål. I afsnit 6 vil den empiriske strategi, model og overvejelser blive uddybet, afsnit 7 præsenterer resultaterne, afsnit 8 undersøger robustheden af resultaterne, afsnit 9 diskutere resultaterne. Afslutningsvis indeholder afsnit 10 konklusionen. Derudover indeholder specialet en litteraturliste og et appendiks.

2 Litteratur

I dette kapitel gennemgås eksisterende forskning, der relaterer sig til mit forskningsspørgsmål. Særligt for både det danske uddannelsessystem og optagelsessystem er, at det er ret unikt, hvilket gør det svært at sammenligne med andre lande. Alligevel inkluderes der i denne litteratur gennemgang internationale studier, hvilket gøres, da de anvendte metoder til at opnå effekter viser sig, at være brugbare for at kunne besvare mit spørgsmål. Der findes en begrænset antal danske studier, som undersøger betydningen af optagelsesknoten for de studerendes frafald, men de som findes beskrives herunder. Slutteligt vil dette afsnit også uddybe, hvordan dette speciale kan bidrage til den eksisterende viden på området.

2.1 Internationale studier

Thistlethwaite og Campbell introducerede i 1960 de grundlæggende idéer bag regression diskontinuitet, RD, designet. En alternativ estimationsmetode, hvor det er muligt at opnå kausale effekter, selvom forskeren ikke selv har kontrol over behandlings- og kontrolgrupper.

Thistlethwaite og Campbell ønskede undersøge, hvad effekterne af college-stipendier betyder for studerende fremtidige præstationer. Med anvendelse af RD designet viser deres resultater, at en offentlig anerkendelse for bedriften på college egnethedsprøver har en tendens til at øge sandsynligheden for, at modtageren modtager et stipendium, men resultaterne tyder ikke på, at anerkendelse påvirker den studerendes holdninger og karriereplaner. (Thistlethwaite & Campbell, 1960)

Denne metode er efterfølgende blevet anvendt til at opnå kausale effekter inden for samfundsøkonomien. Li et al. (2015) har bl.a. anvendt denne fremgangsmåde for at undersøge, hvilken effekt at universitetsstipendier har på de studerendes frafaldssandsynlighed. Ved anvendelsen af denne metode med data fra to italienske universiteter finder de bevis for, at universitetsstipendier er effektive til at forhindre studerende fra familier med lav indkomst i at droppe fra de videregående uddannelser. (Li et al. 2015)

2.2 Danske studier

I en dansk kontekst findes der begrænset litteratur som undersøger betydningen af optagelsesknoten for de studerendes frafaldssandsynlighed.

Tænketaenken DEA udgav i 2014 et notat, som undersøgte om der er forskel i frafaldet for studerende optaget i kvote 1 og kvote 2. I notatet anvendes en simpel sammenligning af kvote 1- og kvote 2-studerende og deres frafald. (DEA, 2014) Deres beregninger viser, at der for 88% af bacheloruddannelserne ikke er signifikante forskelle i frafaldet mellem kvote 1 og kvote 2 ét år efter studiestart. Dog finder de, at på 13 uddannelser klarer de studerende, der er optaget i kvote 2 sig bedre mht. frafald. Disse 13 uddannelser udgør til sammen 20% af det samlede optag af alle bachelorstuderende i perioden 2005-2010, hvilket medfører at det samlede frafald blandt kvote 2-studerende er lavere end kvote 1-studerende. Med deres fund konkluderer de, at synspunktet om, at unge optaget via kvote 2 anses for at være mindre studieegnede end unge optaget via kvote 1 blot er en myte, og dermed er kvote 2 bedre end sit ry. De lægger derfor op til en nuancering af debatten om kvote 2, samt nye måder at optage de studerende på, så det uddannelserne i højere grad får ram på de mest kvalificerede og motiverende ansøgere. (DEA, 2014)

Derudover kortlagde Danmarks Evalueringsinstitution, EVA, i 2019 centrale aspekter af kvote 2-optagesystemet, som anvender de studerendes selvrappede udsagn, om deres optagelse i kvote 2. Denne kortlægning viser, at de studerende optaget i kvote 2 adskiller sig fra de studerende optaget i kvote 1 ved at være mere afklaret og mere motiveret. Disse to faktorer er tidligere undersøgt til, at en væsentlig betydning for at fastholde de studerende. (Danmarks Evalueringsinstitut, 2019)

Retfærdigheden af det danske optagelsessystem er også blevet undersøgt. Et studie af Jens-Peter Thomsen (2016) viser, at motiveret ansøgninger, optagelsesprøver og interviews ikke nødvendigvis bryder den sociale arv, men i stedet medvirker optagelse gennem kvote 2 til en større ulighed i optaget til universiteterne. Han finder, at optagelsessystemet tilgodeser ansøgere fra akademikerhjem ¹, hvilket særligt gør sig gældende i kvote 2. Hvilket forklares af, at ansøgere fra akademikerhjem i højere grad er bedre til at afkode krav og forventninger til at opnå optagelse i kvote 2. Studiet viser også, at ansøgere fra ikke-akademikerhjem har en større chance for at komme ind på en uddannelse i kvote 1 end kvote 2.

Disse undersøgelser viser, at studerende optaget i kvote 2 har i gennemsnit et lavere frafald end studerende optaget i kvote 1. Samtidig problematiseres optagel-

¹ Her er et akademikerhjem karakteriseret ved, at mindst én forælder har en lang videregående uddannelse

sen af ansøgere i kvote 2, da denne medvirker til en større ulighed i optaget til universiteterne. Metoderne anvendt i disse studier er dog simple regressioner eller simple deskriptive sammenligninger. Dermed kan dette speciale bidrage med en mere dybgående analyse af, om der er en effekt af optagelseskvoten på de studerendes frafaldsansynlighed i løbet af det første studieår, og derved give et andet perspektiv på det danske optagesystem.

3 Teori

For at forstå, hvorfor nogle studerende vælger at afbryde deres studie, præsenteres der i dette afsnit forskellige teoretiske frameworks, som kan være med til at belyse dette. Der findes ikke fra et teoretisk perspektiv findes et direkte svar på, hvordan optagelseskvoter påvirker de studerendes frafaldsandsynlighed. I stedet gives der i dette afsnit en kort gennemgang af, human kapital teorien for at forklare, hvorfor den enkelte vælger at uddanne sig. Dernæst gives der mulige forklaringer på, hvorfor studiefrafald finder sted set fra både en økonomisk tilgang og en mere psykologisk-sociologisk tilgang. Indsigten fra disse teoriske rammer sættes efterfølgende i en dansk kontekst, og afslutningsvis indeholder dette afsnit formodede prædikationer og hypoteser, som tager afsæt i den præsenterede teori.

3.1 Hvorfor tage en videregående uddannelse?

I økonomisk litteratur hersker der primært to teoretiske frameworks for at forklare valget om at tage en uddannelse og de medfølgende de senere udfald, som f.eks. afkastet af at have taget en uddannelse. Den ene er human kapital teorien, og den anden er signal teori modellen. Begge disse teorier forudsiger en positiv sammenhæng mellem uddannelse og den enkeltes senere indkomst, hvor den enkelte vil blive aflønnet i forhold til dennes produktivitet.

I human kapital teorien anses den værdi, som individer akkumulerer, såsom viden, evner, social kapital og uddannelse under antagelsen om, at mennesker er produktionsfaktorer på lige fod med fysisk kapital. Dette betyder også, at uddannelse vil give mere human kapital, som øger produktiviteten, og derved vil øge den enkeltes indkomst. Derfor kan valget af uddannelse anses som en investering i human kapital, der giver muligheden for at opnå en højere indkomst på et senere tidspunkt i livet. Dette betyder, at den indkomst, som kunne være opnået under uddannelse, skal ses som en alternativomkostning ved investeringen i at tage en uddannelse. Derfor vil der være et trade-off for den enkelte mellem at uddanne sig, og derved opnå en højere indkomst i fremtiden, og den mulige indkomst, som den enkelte ville kunne opnå nu og her. (Card, 1999) Hvorfor denne teori forudsiger, at et rationelt individ vil derfor vælge at tage en uddannelse, hvis det forventes, at den senere indkomst vil være højere end den nuværende.

I signal teori forstås det derimod, at uddannelse afspejler den enkeltes medfødte

human kapital, og denne medførte human kapital giver anledning til forskelle i produktiviteten og derfor også indkomsten. Produktivitetseffekten af selve uddannelse gør sig således ikke gældende. Uddannelse anvendes i stedet som en mekanisme til at signalere den enkeltes medfødte produktivitet til arbejdsgivere. Individerne med en høj produktivitet vil derved have færre omkostninger ved at tage en lang uddannelse, og derfor vil der foregå en signaleringsmekanisme. (Spence, 1973) Hvorfor denne teori forudsiger, at et rationelt individ med en høj produktivitet vil vælge at tage en uddannelse for at kunne sikre sig en højere fremtidig indkomst.

Disse to økonomiske teorier giver en forklaring på, hvorfor at den enkelte vælger at tage en uddannelse, men giver ikke direkte en forklaring på, hvad optagelseskvoterne har af betydning på de studerende sandsynlighed for at falde fra studiet i løbet af det første studieår. I human kapital teorien kan det argumenteres, at rationelle individer vil afbryde deres studier, hvis de selv anser at deres alternativomkostning er større ved ikke at uddanne sig. Mens det gør sig gældende i signal teorien, at den enkelte vil afbryde deres studie, hvis denne anser, at tage en bestemt uddannelse ikke modsvarer af deres produktivetsniveau og dermed også lønningsmuligheder. Det enkelte individ vil dermed ikke opnå en uddannelse, som modsvarer dennes produktivetsniveau, og derfor vælger den enkelte at afbryde studiet.

3.2 Hvorfor vælger nogle studerende at afbryde deres studie?

Studiefrafaldet dækker over den hændelse, hvor en påbegyndt uddannelse afbrydes inden den enkelte studerende har opnået et bevis på en færdiggjort uddannelse. Dette frafald kan have mange årsager og skyldes ofte et komplekst samspil mellem flere faktorer. Disse årsager kan både handle om manglende trivsel, undervisningens kvalitet, eller at den studerende simpelthen bliver klogere og mere afklaret undervejs.

I en amerikansk kontekst har sociologen Vincent Tinto har udviklet en kendt model om studiefrafald, hvor frafaldet ses som en proces der foregår over tid, og for det andet inddrager den forskellige faktorer som knytter sig til både undervisningen og forhold uden for undervisningen. Centralt for denne teori er, at studiefrafaldet sker i samspillet mellem den enkelte studerende og uddannelsesinstitutioner. Tinto beskriver grundlæggende studiefrafaldet som en frivillig beslutning, og en væsentlig faktor for at kunne forklare frafaldet er studieintegrationen. Ved denne integration forstås sammenhængen mellem den enkelte studerendes engagement og involvering i

uddannelsen, som vil være med til at øge den studerendes tilknytning til uddannelsen. Studieintegrationen i to typer af integration - den sociale integration og den faglige integration. En vellykket integration af begge former vil være med til at fremme den enkelte studerendes tilknytning til uddannelse, hvilket vil være med til at øge engagementet, som vil modvirke studiefrafaldet. Denne integrationsproces afhænger både af individuelle og institutionelle faktorer, som gør sig gældende på forskellig vis i mødet mellem den studerende og den uddannelsen, hvilket giver den enkelte studerende nogle oplevelser, som er med til at påvirke, som denne bliver på studiet eller ej. (Tinto, 1993)

I en dansk kontekst har Studenterrådet ved Aarhus Universitet, 2000, undersøgt dette. Deres undersøgelse viste, at særligt tilfredsheden med det faglige miljø har betydning for de studerendes sandsynlighed til at forlade uddannelse, mens den sociale integration ikke spiller en væsentlig rolle. (Studenterrådet ved Aarhus Universitet, 2000) Andre danske undersøgelser har dog vist, at det sociale miljø, på lige fod med det faglige miljø, har en væsentlig betydning for de studerendes vedholdenhed på de videregående uddannelser. (Danmarks Evalueringsinstitut, 2017) Disse undersøgelser bekræfter dermed, at studiefrafaldet sker i et samspil mellem flere faktorer, som kan være komplekst.

Studiefrafaldet har et tidsperspektiv, som er meget relevant at have for øje, da det tidlige studiefrafald ikke nødvendigvis skyldes de samme faktorer, som det studiefrafald der finder sted senere i uddannelsesforløbet. (Ishitani, 2003) Udvalget for Kvalitet og Relevans på de Videregående uddannelser beskriver frafaldet, som finder sted i løbet af det første studieår som en indikation på, at der har fundet en dårlig forventningsafstemning mellem den enkelte studerende og uddannelsen, og derfor har den enkelte studerende ikke ved studiestarten haft en viden stor nok til at vide, hvad denne gik ind til. (Kvalitetsudvalget, 2015). Dette speciale undersøger netop det tidlige studiefrafald, og dermed kan de senere opnåede resultater af den empiriske analyse udelukkende fortolkes for det frafald der sker inden for det første studieår.

3.3 Teoretisk hypotese

Den beskrevne teori ovenfor giver anledning til at opstille følgende hypotese for dette speciale:

Studerende, som bliver optaget i kvote 2, er mindre tilbøjelig til at falde fra i løbet af det første studieår

Denne hypotese kan i kombination af den præsenterede teori ovenfor forklares med, at de studerende, som optages i kvote 2 har ydet en ekstra indsats for at blive optaget i form af f.eks. en motiveret ansøgning. Det betyder, at den enkelte studerende allerede tidligere i ansøgningsprocessen har gjort op med sig selv, om denne ekstra indsats på sigt vil kunne svare sig i forhold til den enkeltes alternativ omkostning eller om ansøgte uddannelse er det rigtige signal til at signalere den enkeltes medførte produktivitet. Derudover må det formodes, at den enkelte studerende med optagelse i kvote 2 har et større engagement og involvering i uddannelse sammenlignet med studerende, som er optaget i kvote 1, da optagne i kvote 2 har ydet denne ekstra indsats for at blive optaget. Dette speciale vil forsøge, at enten be- eller afkræfte denne hypotese.

4 Det danske optagesystem

I dette afsnit gennemgås kendetegnene ved det danske optagesystem til de videregående uddannelser. Der har gennem flere årtier været et politisk ønske om, at fremtidige generationer skal være mere uddannet end den forrige, hvilket har medført en markant stigning i antallet af studerende på de videregående uddannelser, hvor der årligt optages cirka 90.000 studerende. Fælles for alle uddannelser er, at ansøgerne skal have en gymnasial eksamen eller en anden adgangsgivende uddannelse og have taget bestemte gymnasiale fag på bestemte niveauer evt. med en minimumskaraktter. (Uddannelse- og forskningsministeriet, 2020)

I Danmark er næsten alt uddannelse, herunder også en videregående uddannelse, gratis. Studerende på en videregående uddannelse modtager desuden også en månedlig betaling på omtrent 6.000 kroner. For udeboende studerende på en videregående uddannelse⁰, samt der er mulighed for at låne yderligere penge hos staten i kombination med SU'en. Dette gør det særligt attraktivt for den enkelte at uddanne sig. Den gratis uddannelse stilles til rådighed af staten, idet at en uddannet befolkning er med til at sikre velstand, hvorfor det også er attraktivt for samfundet, at have en veluddannet befolkning, som tidligere nævnt i afsnit 3.

4.1 Den Koordinerede Tilmelding

Alle ansøgninger til de videregående uddannelser i Danmark håndteres af et centraliseret optagesystem, som også er kaldt den Koordinerede Tilmelding (KOT). Når en ansøger vælger, at søge uddannelse har hver ansøger har muligheden for at søge op til 8 forskellige videregående uddannelser i den samme ansøgning. Det betyder også, at systemet hvert år håndterer mange ansøgninger, som skal fordeles på de udbudte studiepladser. I 2018 håndterede KOT 241.826 ansøgninger til de videregående uddannelser, som blev fordelt på 64.943 studiepladser. (Uddannelses- og forskningsministeriet, 2018) For størstedelen af ansøgningerne gør det sig gældende, at på de fleste uddannelser er der flere ansøgninger end der er pladser, hvilket medfører at ansøgerne muligvis er begrænset i deres valg af uddannelse.

KOT fordeler ansøgerne til uddannelserne således, at de bedste ansøgere bliver tildelt en studieplads på deres fortrukne uddannelser. Systemet er kompleks, men størstedelen af ansøgerne fordeles udelukkende på baggrund af deres adgangsgivende karakter gennemsnit fra en ungdomsuddannelse. Derudover har de fleste uddan-

nelser nogle simple restriktioner om bestemte fag fra en ungdomsuddannelse. Hver uddannelsesinstitution indberetter til KOT, hvor mange studiepladser de har på de enkelte uddannelser.

Mange uddannelser anvender simple optagelsesregler for ansøgere med et karaktersnit, som præcis er adgangskvotienten; de tilbyder alle ansøgere en plads, eller an anden anvendt regel er en aldersregel, som medfører at hvis to ansøgere er har det samme karaktersnit som svarer til adgangskvotienten tilbydes den ældste ansøger pladsen. For andre uddannelser benyttes andre optagelsesregler for marginal ansøgere. (Uddannelses- og forskningsministeriet, 2018)

4.1.1 Kvotesystemet

Det danske optagelsessystem er derudover karakteriseret ved at være et kvotesystem, hvilket betyder, at der gøres brug af kvoter på de uddannelser, hvor der er flere ansøgere end der er studiepladser. Således vil studiepladserne blive fordelt gennem to kvoter, kvote 1 og kvote 2.

Udvælgelsen i kvote 1 foretages udelukkende på baggrund af ansøgernes adgangsgivende karaktergennemsnit. Ansøgerne er rangeret efter deres adgangsgivende karaktergennemsnit fra en ungdomsuddannelse. Ansøgere med det højeste karaktergennemsnit bliver rangeret højest, og vil få tildelt en studieplads, hvorefter græsekvotienten for den bestemte uddannelse bliver kvotienten på den marginale ansøger. Det betyder, at ansøgere med et karaktergennemsnit over græsekvotienten bliver tilbudt en plads, mens ansøgere med et karaktersnit lavere end græsekvotienten er afvist. Dette medfører, at græsekvotienten til den enkelte uddannelse bestemmes hvert år efter ansøgningsfristen. Selvom at antallet af studiepladser kun varierer ganske lidt år til år, så varierer antallet af ansøgere, hvilket betyder at græsekvotienten varierer. Ansøgerne kender derfor ikke græsekvotienten på ansøgningstidspunktet, hvorfor det vil ikke være muligt for den enkelte ansøger at forudsige om græsekvotienten vil være over eller under ansøgerens karaktergennemsnit, hvilket særligt gør sig gældende for de ansøgere, som har et karaktergennemsnit tæt på græsekvotienten. (Uddannelses- og forskningsministeriet, 2018)

I kvote 2 sker udvælgelse med inklusion af andre kriterier end karaktergennemsnittet. I kvote 2 har uddannelserne mulighed for at tildele studiepladserne efter en

konkret vurdering og rangering af ansøgere. Uddannelserne kan med denne kvote forsøge at få informationer om ansøgernes generelle og specifikke faglige forudsætninger og motivation for uddannelsen, end hvad de ellers ville have opnået med ansøgernes karaktergennemsnit. Formålet med kvote 2 er dermed at give ansøgere muligheden for en studieplads, hvis de ikke har et tilstrækkeligt karaktergennemsnit.

Det er op til de enkelte uddannelser at fastlægge kvotefordelingen, men på universiteterne har størstedelen af uddannelserne en kvotefordeling med 90% af pladserne afsat til kvote 1, mens 10% af pladser er afsat til kvote 2. Der er dog forskelle blandt de danske universiteter, hvor særligt Syddansk Universitet og Roskilde Universitet er de universiteter med den største andel af studiepladser forbeholdt til kvote 2.

4.2 Den sociale mobilitet i optagesystemet

Det danske optagesystem forsøger, at tilgodese en masse hensyn, således at optaget på de videregående uddannelser sker på de bedste vilkår, som både bidrager til at mindske frafaldet og dimittendledigheden, samt sikre et højt fagligt niveau på uddannelserne. Det danske optagesystem forsøger at give muligheden for, at alle kvalificerede ansøgere kan søge optagelse ubetinget af forældres uddannelsesniveau. Derved forsøger man politisk at øge uddannelsesmobiliteten. Sandsynligheden for at blive optaget på en videregående uddannelse er nemlig markant lavere for unge med faglærte og ufaglærte forældre sammenlignet med unge med forældre med en lang videregående uddannelse. (Uddannelses- og forskningsministeriet, 2020)

Elementet af den sociale mobilitet i optagesystemet er også en del af forklaringen bag indretningen af kvotesystemet, da kvote 2 til tider anses som en 'second chance' for de studerende, som ikke har opnået et karaktergennemsnit fra en ungdomsuddannelse, som opfylder grænsekvotienten på den ønskede uddannelse. Disse studerende skal således ikke begrænses i deres uddannelsesvalg, hvis ambitionerne for en bestemt uddannelse er til stede. Ansøgere som ikke har opnået et karaktergennemsnit, som lever op til grænsekvotienten vil have muligheden for at søge om optagelse til uddannelsen gennem kvote 2 i stedet.

Tidligere undersøgelser har vist, at variationen i de studerendes karaktergennemsnit fra en ungdomsuddannelse til en vis grad kan forklares af forældres uddannelsesniveau, hvorfor forældres uddannelsesbaggrund spiller en rolle for de studerendes mulighed. (EVA, 2015) Ansøgere med forældre med en videregående uddannelse får

højere karakter på de gymnasiale uddannelser, hvorfor disse vil være bedre stillet i kvote 1. Dette betyder, at ansøgere med faglærte eller ufaglærte forældre er ringere stillet i kvote 1. Derudover er unge med faglærte eller ufaglærte forældre mindre tilbøjelige til at gennemføre en ungdomsuddannelse, og derved også søge ind på en videregående uddannelse end unge med forældre med en videregående uddannelse.

Derved begrænses mulighederne for at optagesystemet kan være med til at påvirke den sociale mobilitet, da den forekommer tidligere i uddannelsessystemet. Selvom optagesystemet forsøger at tage hensyn til den sociale mobilitet, kan dette være svært, da en stor del af udskillelsen allerede er sket på et tidligere tidspunkt.

Væsentlig er det også, at med de anvendte udvælgelseskriter i kvote 2 som f.eks. en motiveret ansøgning vil ansøgere som har forældre med en videregående uddannelse klare sig bedre og i højere grad opnå en optagelse i kvote 2 sammenlignet med en ansøger med faglærte eller ufaglærte forældre. Dette kan forklares med, at ansøgere med forældre med en videregående uddannelse er mere tilbøjelig til at vide, hvad der skal stå i denne ansøgning, og de kender det faglige kodeks. Det gør sig netop gældende, at for at kunne søge ind på en videregående uddannelse, skal ansøgerne kunne forstå optagesystemet. De skal vide, hvor de kan søge information om uddannelserne, hvad de forskellige begreber som f.eks. adgangskrav og grænsekvotienter betyder, og hvad der skal til for, at de kan kvalificere sig til at blive optaget på en videregående uddannelse. Den viden er afgørende for ansøgernes mulighed for at foretage en kvalificeret ansøgning og i sidste ende at blive optaget. Ansøgere som har forældre med en lang videregående uddannelse vil i højere grad have en bedre forståelse for optagesystemet, hvorfor disse vil være bedre stillet sammenlignet med ansøgere, som har forældre med en faglært eller ufaglært uddannelsesbaggrund. (Thomsen, 2014)

5 Data

Dette afsnit vil beskrive, hvilke data der anvendes til den empiriske analyse. Det primære datagrundlag er fra danske administrative registre, som administreres af Danmarks Statistik. Dataadgang er opnået gennem adgang via Danmarks Evalueringsinstitut. I afsnittet præsenteres de anvendte registre og variabler, som der er konstrueret til analysen. Derefter gennemgås udvælgelsen af populationen, og herefter præsenteres relevante beskrivende deskriptive statistikker.

Tabel 5.1: Data kilder

Register	Data
BEF	Karakteristika om bl.a. køn, alder, oprindelsesland, samt forældre
IND	Forældres indkomst ved barnets 15. leveår
INST	Uddannelsesinstitution karakteristika
KOT	Informationer om den koordinerede tilmelding herunder optagelsekvote og adgangsgivende karaktergennemsnit
KOTRE	Påbegyndte og afsluttede danske uddannelser
UUDA	Forældres højest fuldførte uddannelse ved barnets 15. leveår
Andet	
Grænsekvote	Årlige grænsekvote på danske videregående uddannelser
Kvote 2 kriterier	Kriterier som anvendes for optagelse i kvote 2
Afklarethed	Selvrapporteret niveau af afklarethed om studievalg

Note: Grænsekvote og Kvote 2 kriterier er ikke registre, men i stedet indhentet via Uddannelse- og forskningsministeriet. Mens information om de studerendes afklarethed er opnået via data fra Danmarks Evalueringsinstitut

5.1 Data kilder og variable

Data, som er anvendes til den empiriske analyse, er administrative registerdata fra Danmarks Statistik og indeholder informationer om alle individer, der i perioden 2016-2018 har søgt en videregående uddannelse. Hvert individ har et unikt identifikationsnummer, hvilket gør det muligt, at kombinere informationer fra flere registre og følge disse individer over tid. De primære registre har en årlig opdateringsfrekvens.

Jeg kombinerer registre med deres tilhørende tilgængelige informationer, som er relevant for at udføre den empiriske analyse, som kan ses i tabel 5.1. Derudover anvendes også data om de studerendes selvrapporterede afklarethed, som er opnået ved brug af Danmarks Evalueringsinstituts forløbsundersøgelse. Denne forløbsundersøgelse udsendes til alle ansøgere på de danske videregående uddannelser. Informationer om de årlige grænsekvotienter og kvote 2 kriterier er ligeledes indhentet, som vil være særlig relevante for den empiriske analyse. Nedenfor vil de enkelte datakilder blive gennemgået yderligere.

5.1.1 Den Koordinerede Tilmelding

Det primære dataregister til den empiriske analyse er KOT-registret, som administreres af Danmarks Statistik og indeholder informationer om optagne på de videregående uddannelser og deres ansøgning til denne optagelse. Data er opgjort med en årlig opdateringsfrekvens, således at hvert år indeholder informationer om ansøgninger og optagne i det pågældende år. Dette datasæt indeholder desuden detaljerende oplysninger om, hvilke optagelseskvote som den enkelte er optaget i, ansøgnernes karaktergennemsnit, samt informationer om de ansøgers prioritering af uddannelser.

Som nævnt i afsnit 4 er det muligt, at hver ansøger kan ansøge op til 8 uddannelser hvert år i en prioriteret rækkefølge. I dette datasæt optræder et individ hver gang vedkommende har lavet en ansøgning med en prioritetsliste til de videregående uddannelser. I data fremgår ansøgerens prioritetsnummer fra 1-8, en unik uddannelseskode for hver søgt uddannelse. I datasættet findes der ligeledes informationer om, hvordan den enkelte ansøgning er håndteret af KOT. Her er det muligt, at ansøgningen har ført til optagelse, en afvisning, en clearing eller en afmeldelse. En clearing betyder, at den enkelte ansøger er optaget på en højere prioritet, mens en afmeldelse betyder, at den enkelte ansøger selv har afmeldt sig eller ikke lever op til de eventuelle fagspecifikke krav. Ansøgninger der er clearet eller afmeldt vil blive sorteret fra inden at tildelingsprocessen for kvote 1 ansøgninger finder sted, hvorfor disse ansøgninger aldrig vil være afprøvet på deres karaktergennemsnit fra en ungdomsuddannelse. Dette betyder, at hvis en ansøger har søgt om optagelse på jura som sin 1. prioritet og opfylder kravene, så vil denne blive tilbudt en plads, men vil optræde som 'clearet' i de resterende ansøgninger, da denne allerede har modtaget

en studieplads på en højere prioritet. Omvendt vil en ansøger, som har søgt optagelse på jura, men ikke har eksamen fra et adgangsgivende ungdomsuddannelse, eller ikke lever op til et fagspecifikt krav få koden som værende 'frameldt'.

Derudover indeholder datasætterne informationer om ansøgerens karaktergennemsnit, som det fremgår fra indberetning fra den pågældende ungdomsuddannelse. Dette karaktergennemsnit indeholder evt. bonusser, som har været gældende for ansøgningen. Disse bonusser kan være opnået, hvis den pågældende har haft ekstra fag på A-niveau², eller er startet på en videregående uddannelse senest to år efter afsluttet ungdomsuddannelse³.

Registret indeholder også informationer om, hvilken optagelseskvote at den studerende er optaget i. Hver kvote indeholder muligheden for optagelse enten direkte eller på et ventelistesystem (standby), hvilket betyder, at der vil være mulighed for fire forskellige søgemuligheder. Hver ansøger vælger en af disse muligheder, når de søger uddannelse. Tildelingen af studiepladser fungerer på den måde, at alle ansøgere først prøves i kvote 1 uanset, hvilken kvote at den enkelte ansøger har søgt. Hvis ansøgeren opfylder kravene for kvote 1, så bliver denne ansøger tilbudt en studieplads i kvote 1, og dermed vil denne ansøger fremgå som en kvote 1 ansøger i optagelse. Det betyder, at en ansøger, som har søgt en kvote 1 standby plads med et karaktergennemsnit fra en ungdomsuddannelse, som har et højt nok snit til at komme ind på kvote 1, vil derfor blive optaget som en kvote 1 ansøger. Derved vil der opstå en forskel i mellem den søgemulighed og optagelsen. Tilsvarende vil det også være muligt at en kvote 2 ansøger først vil blive prøvet i kvote 1, dernæst i kvote 1 standby og så kvote 2 og til sidst i kvote 2 standby. Dette giver fire muligheder for optagelse, og dermed kan den enkelte ansøger blive registreret som optaget på en af de fire.

Derudover er det væsentlig at nævne, at registreringen af ansøgere optaget på standby vil have en pladsgaranti det efterfølgende år. Dette betyder, at det ikke er muligt, at sammenligne på samme vis hvilken grænsekotient, at disse er optaget på. Denne gruppe af studerende, som er optaget på standby pladser udgør dermed en udfordring, selvom det vurderes at disse udgør en relativ lille andel af det samlede

² Denne bonus muliggør det, at gange karaktersnittet med 1,03 for et ekstra A-fag og 1,06 for to ekstra A-fag

³ Hurtigstartsbonusen blev fjernet i 2020, men vil fortsat være relevant i indeværende analyse, da datagrundlaget er i perioden 2016-2018. Denne bonus muliggjorde det, at gange karakter-snittet med 1,08

optag. Håndteringen af denne gruppe vil blive belyst i afsnit 5.2

5.1.2 Elevregistret

For at besvare, hvordan optagelseskvoten kan påvirke universitetsbachelors frafaldssandsynlighed, er variablen for frafald væsentlig. Denne information opnås gennem elevregistret, KOTRE, som indeholder information om påbegyndte og afsluttede uddannelser, herunder også folkeskole, ungdomsuddannelser, videregående uddannelser og efteruddannelsesforløb. Data er opgjort med en årlig opdateringsfrekvens, således at hvert år indeholder informationer om påbegyndte, igangværende og afsluttede uddannelsesforløb.

Elevregistret muliggør det at opnå informationer om de studerendes frafald. Dette gøres ved at konstruere en frafaldsvariabel, som gøres ved brug af uddannelseskvalifikationskode, som en uddannelse kan føre til, samt datoen for den påbegyndte uddannelse. Hvis den enkelte studerende har afbrudt sin uddannelse indenfor det første studieår, som er målt fra den 1. oktober i det pågældende år til 1. august i det efterfølgende år, og ikke har opnået en uddannelseskvalifikation, så vil denne studerende blive karakteriseret som værende frafaldet.

Denne variabel er meget væsentlig for den empiriske analyse, at denne udgør udfaldsvariablen. Variablen er konstrueret som værende en dummy-variabel, som antager værdien 1, hvis den studerende er faldet fra i løbet af det første studieår, og 0 hvis den studerende fortsat læser den pågældende uddannelse.

5.1.3 Grænsekvotienter og kvote 2 kriterier

En samlet oversigt over grænsekvotienterne til universitetsbachelor i perioden 2016-2018 er fremsendt af Uddannelses- og Forskningsministeriet. Data indeholder informationer om alle grænsekvotienter på de videregående uddannelser i Danmark. Mange af disse uddannelser vil også være angivet uden en grænsekvotient, hvorfor de ikke vil være relevante for analysen, hvilket beskrives nærmere i afsnit 5.2. Grænsekvotienterne er væsentlige i den empiriske analyse, da de vil anvendes som en del af afstandsvariablen, hvilket beskrives nærmere i afsnit 6.2.

Derudover er der ligeledes fremsendt af Uddannelses- og forskningsministeriet informationer, hvilke kriterier der lagde til grund for optagelse i kvote 2 for det

enkelte uddannelsesudbud. Disse er karakteriseret ved tre kategorier:

- *Fremmøde*

Dette udvælgelseskriterie dækker over en samtale eller test har været anvendt for at teste om ansøgeren skal optages i kvote 2. I kombination med disse er det muligt, at ansøgeren også skal have vedhæftet en motiveret ansøgning.

- *Passiv uden motiveret ansøgning*

For dette udvælgelseskriterie gør det sig gældende, at vurderingen af ansøgeren har fundet sted på baggrund af baggrund af fagspecifikke karakter i kombination af karaktergennemsnittet fra en ungdomsuddannelse. Det er meget få uddannelser, som anvender dette udvælgelseskriterie.

- *Passiv med motiveret ansøgning*

I dette udvælgelseskriterie gør det sig gældende, at en motiveret ansøgning har været primær udvælgelse for optagelse i kvote 2. Udvalgeskriteriet kan dog også være i kombination af den motiveret ansøgning sammen med karaktergennemsnit fra en ungdomsuddannelse. Dette kvote 2 kriterie er det mest anvendte.

5.1.4 Kontrolvariable

Datasættene fra Danmarks Statistik er administrative registerdata, som indeholder detaljeret information om hele den danske befolkning. Fra disse registre er det muligt, at opnå oplysninger om blandt andet indkomst, uddannelse og familiebaggrund. For at konstruere et datasæt, som er relevant for den empiriske analyse, kombineres KOT-datasættet med individuelle oplysninger, som fås gennem de oplyste registre, som præsenteret i tabel 5.1. Kombinationen sker ved hjælp af en unik identifikator for hver enkelt individ. Derved opnås der et kombineret datasæt, som indeholder informationer om optagelseskvote til de videregående uddannelser, uddannelsesvalg og baggrundsvariabler, som bl.a. dækker over køn, alder, herkomst, forældres indkomst, forældres højeste uddannelsesniveau.

Informationer om individers baggrundskarakteristika fås fra befolkningsregisret, BEF. Dette register indeholder data om den enkelte optagnes køn, alder, oprindelsesland. Disse informationer er væsentlig for den senere empiriske analyse, da disse vil indgå som kontrolvariable. Derudover opnås der også informationer, om forældres

højeste fuldførste uddannelsesniveau gennem registret, UDDA. Dette register indeholder informationer om forældres højeste fuldførte uddannelsesniveau, som er målt ved barnets 15. leveår. Ligeledes opnås der information om forældrenes indkomst via indkomstregistret, IND. Denne indkomst er også målt ved barnets 15. leveår. Da populationen af optagne ikke var 15 år på samme tidspunkt vil denne variabel være inflateret til 2010-niveau, således at den er mulig at sammenligne på tværs af optagne studerende. Derudover vil indkomstvariablen være log-transformeret

5.1.5 Spørgeskemadata

For at nuancere og undersøge, hvor robuste resultaterne er i analysen inkluderes der et mål for den enkeltes niveau af afklarethed. Denne information er opnået ved brug af en spørgeskemaundersøgelse, som Danmarks Evalueringsinstitut, EVA, har udsendt til alle studerende, som er blevet tilbudt en studieplads på en videregående uddannelse siden 2016. Spørgeskemainsamlingerne laves i flere bølger i løbet af et studieår. I denne analyse vil data fra den første bølge blive benyttet, som udsendes i august måned, hvorfor de studerende endnu ikke er startet på studie endnu. Spørgsmålet, som bruges som et mål for de studerendes afklaret, lyder således:

"Da jeg søgte, følte jeg mig helt afklaret med, at jeg ville studere på netop den uddannelse, jeg nu er blevet optaget på"

Spørgsmålet er en besvares med en skala af svar, hvor *"slet ikke"* angiver det laveste niveau og *"I høj grad"* angiver det højeste niveau af afklarethed. Det er udelukkende muligt, at opnå informationer om de studerende, som aktivt selv har valgt af svare på spørgeskemaet. Dette kan medføre en vis skævhed sammenlignet med den fulde population af de studerende på de videregående uddannelser. Derfor giver appendiks A tabel .1 et overblik over, hvor repræsentativ stikprøven af respondenter som har besvaret spørgeskemaet er sammenlignet med den fulde population. Det fulde overblik over indholdet i dette spørgeskema findes ligeledes i appendiks A.

5.1.6 Variableoversigt

Ovennævnte registre og datakilder medfører, at Tabel 5.2 viser, hvordan de anvendte variable er konstrueret, samt deres datakilde. En mere beskrivende gennem-

gang af variablene kan ses i appendiks B.

Tabel 5.2: Variabeloversigt

	Variabel	Type	Kilde
Udfaldsvariabel	Førsteårsfrafald	Dummy	KOTRE
Behandlingsvariabel	Adgangsgivende karaktergennemsnit	Kontinuert	KOT
Kontrolvariable	Køn	Dummy	BEF
	Alder	Kontinuert	BEF
	Herkomst	Kategorisk	BEF
	1. prioritet	Dummy	KOTRE
	Uddannelse	Kategorisk	KOT
	Akademisk uddannet mor*	Dummy	UDDA
	Akademisk uddannet far*	Dummy	UDDA
	Mor log indkomst*	Kontinuert	IND
	Far log indkomst*	Kontinuert	IND

Note: Adgangskvotetienter og Kvote 2 kriterier er ikke registre, men data er fremskaffet efter kontakt med Uddannelse- og forskningsministeriet. *Forældres uddannelse og indkomst er målt i den studerende 15. leveår

5.2 Data restriktioner og population

Grundlaget for den empiriske analyse er studerende, som er optaget på en universitetsbachelor gennem den Koordinerede Tilmelding ved den ordinære tilmelding med ansøgningstidsfrister i marts og juli. Derfor er informationer om optagne studerende i perioden 2016-2018 nødvendige. Ansøgninger som ikke har ført til optagelse anvendes i dette speciale. Denne afgrænsning foretages, da jeg udelukkende ønsker at undersøge, hvilke effekt optagelseskvoten har på de studerendes frafaldssandsynlighed. Ansøgere, som ikke er optaget, vil derved ikke have en optagelseskvote, og derfor overvejes disse ikke. Derudover foretages der yderligere restriktioner for at opnå et homogent datasæt, hvor den eksogene variation udelukkende stammer fra optagelseskvoten, hvorfor det er nødvendigt at foretage restriktioner i data.

Først og fremmest afgrænses data til kun, at indeholde informationer om indi-

vider optaget på en universitetsbachelor. Det medfører, at studerende optaget på professionsskoler og erhvervsakademier frasorteres. Med denne frasortering sikres det, at uddannelseslængden ikke varierer på tværs af uddannelsessektorerne, hvilket kan være vigtigt, da det er plausibelt at uddannelseslængden har betydning for, hvornår i studiet at de studerende falder fra. Datasættet indeholder informationer om cirka 70.000 optagne universitetsbachelor i perioden 2016-2018. Tabel 5.3 viser de efterfølgende datarestriktioner.

Tabel 5.3: Data restriktioner

Beskrivelse	Antal observationer	Andel i pct.
Optaget på en studieplads på et universitet i 2016-2018	69.769	100,0
Optaget på en studieplads med frit optag	- 14.741	21,1
Optaget på en studieplads med udelukkende kvote 2-optag	- 1.362	2,0
Optaget på en studieplads uden et karaktersnit fra ungdomsuddannelse	- 280	0,4
Optaget på en studieplads via efteroptag eller standby-plads	- 2.363	3,4
I alt	51.023	73,1

Kilde: Egne beregninger baseret på registerdata fra Danmarks Statistik

Formålet i dette speciale er, at undersøge effekten af optagelseskvoten på frafaldsandsynligheden. Derfor er det ikke relevant, at inkludere uddannelser hvor der ikke opereres med flere optagelseskvoter. Observationer fjernes derfor, hvis de studerende er optaget på en studieplads, hvor der i det pågældende år har været frit optag eller ingen grænseskvotient. Dette er et krav for den empiriske analyse, da en fraværende grænseskvotient vil medføre alle ansøgere vil være optaget i kvote 1.

Det samme gør sig gældende, hvis en uddannelse udelukkende anvender optagelse i kvote 2. Det vil betyde, at der på den pågældende uddannelse kun opereres med én optagelseskvote, og derfor vil denne uddannelse ikke være relevant for denne analyse. Disse restriktioner om, at uddannelser skal have to optagelseskvoter medfører en mindskelse på omtrent 23% af de samlede observationer, som det kan ses i tabel 5.3. Det bemærkes, at størstedelen af universitetsuddannelser gør brug af flere optagelseskvoter.

Et andet centralt element i den empiriske strategi er de studerendes adgangsgivende karaktergennemsnit. Denne information skal bruges til at beregne afstanden

mellem den studerendes karaktergennemsnit og grænsekvisitienten for den pågældende uddannelse. Optagne som ikke har et adgangsgivende karaktergennemsnit, vil derfor blive frasorteret. Individer uden et karaktersnit skyldes primært, at de har en udenlandsk ungdomsuddannelse eller en anden kvalificerende uddannelse, hvor der ikke er registreret et karaktergennemsnit. Det er derved ikke muligt, at indhente informationer for disse studerende, hvorfor de fjernes fra datasættet.

Desuden vil studerende, som er optaget gennem efteroptag eller studerende, som er optaget direkte på de videregående uddannelseinstitutioner om vinteren fjernes. Studerende, der har fået en standby-plads i det forrige, er ligeledes ekskluderet fra analysen. For en studerende som er optaget via efteroptag eller på en standby-plads er det muligt, at grænsekvisitienten ikke er sammenligning med den ordinære, hvilket kan skabe uoverensstemmelse, hvorfor disse observationer fjernes, for at opnå så homogent et datasæt som muligt.

Ovennævnte restriktioner medfører, at datasættet består af 51.023 observationer. Disse observationer inkluderer alle optagne på en universitetsbachelor i perioden 2016-2018, som overholder ovenstående restriktioner. Der findes i appendiks A en balancetabel, som viser forskellen mellem det restrikeret datasæt og den fulde population på række interessevariable.

5.3 Deskriptiv statistik

I dette afsnit udføres der en deskriptiv analyse af henholdsvis de optagne studerende i kvote 1 og kvote 2, samt den fulde population. Den deskriptive statistik kan ses i tabel 5.4.

I denne tabel kan det ses, at kvinder udgør størstedelen af de optagne studerende, hvilket både gør sig gældende i kvote 1 og kvote 2. Dog er kvindeandelen i kvote 2 4 pct. point mindre sammenlignet med kvote 1. Dette betyder, at en større andel af mænd udgør optaget i kvote 2 sammenlignet med kvote 1. Denne forskel kan muligvis forklares med, at kvinder i gennemsnit får højere karakter på ungdomsuddannelser end mænd, hvilket kan være med til, at medføre at mænd i højere grad anvender kvote 2 for optagelse til en videregående uddannelse. (Dansk Erhverv, 2019)

Overraskende er det dog ikke, at de optagne studerende i kvote 2 er ældre end de optagne i kvote 1. Dette kan skyldes, at der i kvote 2 findes kriterier for optagelse i denne kvote, som beskrevet i 5.1.3. Et kriterie kan være at ansøgeren eksempelvis

skal have relevant erhvervserfaring eller været på et udlandsophold for at skrive en motiveret ansøgning. Dette medfører, alt andet lige, at kvote 2 ansøgere i mindre grad søger direkte videre på en videregående uddannelse efter færdiggørelse af en ungdomsuddannelse, da de først herefter skal ud opnå erhvervserfaringen eller udlandsopholdet. Derfor vil det også være naturligt, at gennemsnitsalderen i kvote 1 være lavere end kvote 2.

Tabel 5.4: Deskriptiv statistik

	Kvote 1			Kvote 2			Total		
	Gennemsnit	Std. fejl	Antal	Gennemsnit	Std. fejl	Antal	Gennemsnit	Std. fejl	Antal
	observationer			observationer			observationer		
Kvinde	0,55	0,50	43.185	0,51	0,50	7.816	0,54	0,49	51.023
Alder	21,45	3,45	43.185	22,83	4,15	7.816	21,53	3,78	51.023
Dansk herkomst	0,88	0,33	43.185	0,85	0,37	7.816	0,88	0,32	51.023
STX adgangsgivende eksamen	0,56	0,49	43.185	0,60	0,49	7.816	0,54	0,49	51.023
Adgangsgivende karaktergennemsnit	9,70	1,83	43.185	7,10	1,94	7.816	9,44	1,84	51.023
1. prioritet	0,84	0,36	43.185	0,74	0,44	7.816	0,83	0,78	51.023
Akademisk uddannet mor	0,22	0,41	43.185	0,20	0,40	7.816	0,21	0,41	51.023
Akademisk uddannet far	0,29	0,45	43.185	0,29	0,45	7.816	0,29	0,45	51.023
Mors indkomst (log)	12,81	0,62	43.185	12,76	0,73	7.816	12,81	0,63	51.023
Fars indkomst (log)	13,06	0,85	43.185	13,01	0,97	7.816	13,06	0,86	51.023
Førsteårsfrafald	0,19	0,39	43.185	0,13	0,34	7.816	0,18	0,38	51.023
Selvrapporteret afklarethed	0,35	0,48	22.295	0,49	0,50	3.749	0,37	0,49	26.044

Note: Tabellen viser middelværdierne for de tre grupper: studerende optaget i kvote 1, studerende optaget i kvote 2 og det samlede optag af studerende i perioden 2016-2018.

Kilde: Egne beregninger baseret på registerdata fra Danmarks Statistik

Fra tabellen kan det ligeledes aflæses, at de optagne studerende i kvote 1 er i højere grad optaget på deres 1. prioritet sammenlignet med kvote 2. Derudover ser det ikke ud til, at kvote 2 tilgodeser individer med forældre med en lang videregående uddannelse. Andelen af optagne studerende som har en mor eller en far med en lang videregående uddannelse i kvote 1 og kvote 2 ser ud til at være nogenlunde lige store. Dog ser det ud til, at der er en lille forskel på forældrenes indkomst for optagne i kvote 1 sammenlignet med kvote 2. Studerende optaget i kvote 1 har forældre med en højere indkomst end studerende optaget i kvote 2. Denne forskel er dog af en begrænset størrelse.

Der kan derudover bemærkes, at studerende som er optaget i kvote 2 oplyser selv, at de er mere afklaret med deres uddannelsesvalg sammenlignet med studerende, som er optaget i kvote 1. Denne information er opnået gennem spørgeskemadata, som beskrevet i afsnit 5.1.5. Dette virker også plausibelt alene af den grund, at de studerende i kvote 2 senest indleverer deres ansøgning 3 måneder før den seneste frist for ansøgning i kvote 1.

Sammenligning af de to gruppers gennemsnitlig frafald understøtter hypotesen om, at studerende optaget i kvote 2 tenderer til at falde mindre fra end studerende optaget i kvote 1. Det kan aflæses af tabel 5.4, at studerende optaget i kvote 2 falder gennemsnitlig 6 pct. point mindre fra sammenlignet med studerende optaget i kvote 1. Denne forskel i frafaldet kan dog ikke fortolkes kausalt på grund af endogenitet. Men det ser ud til, at der er en forskel for de to gruppers frafaldssandsynlighed.

6 Empirisk strategi

I dette afsnit introduceres den empiriske strategi, som anvendes til at besvare mit forskningsspørgsmål. Målet er at estimere den kausale effekt af optagelseskvoten for de studerendes frafaldssandsynlighed. Som nævnt i tidligere afsnit 4, kan de industrielle rammer udnyttes i et regression diskontinuitet design, som giver muligheden for at identificere og estimere den kausale effekt af optagelseskvotens betydning for de studerendes frafaldssandsynlighed. I designet udnyttes det faktum, at sandsynligheden for optagelse ændres diskontinuert som en funktion af grænsekvotienten og ansøgerens karaktergennemsnit fra en ungdomsuddannelse. Grænsekvotienterne vil benyttes som en sorteringsmekanisme, som bestemmer hvorvidt et individ bliver optaget eller ej. Dette medfører at grænsekvotienterne anvendes til at definere en cut-off værdi for hver uddannelse hvert enkelt år. Den underliggende identifikation i designet er, at ansøgere lige over og under denne grænsekvotient i alle væsentlige henseender er ens. Første del af afsnittet vil gennemgå de fundamentale idéer bag regression diskontinuitet designet. Hernæst fremlægges den økonometrisk specifikation og slutteligt introduceres en alternativ estimationsmetode for at undersøge, om det er muligt at minimere eventuelle endogenitetsudfordringen.

6.1 Regression diskontinuitet design

Et stort ønske i samfundsvidenskaben er, at kunne undersøge og forklare de kausale effekter af politiske interventioner. Ligeledes ønsker jeg med mit forskningsspørgsmål, at besvare hvad den kausale effekt af behandlingsvariablen, optagelseskvoten, er på den afhængige variabel, frafaldssandsynligheden. Undersøgelser af disse effekter er umiddelbart også lige til, når det er tilfældigt, hvem der modtager sådan en behandling. Den tilfældige tildeling ville netop medføre, at en simpel OLS regression ville give den gennemsnitlige treatment-effekt (ATE) af optagelseskvoten på de studerendes frafaldssandsynlighed betinget på deres baggrundskaraktistika. I praksis er det dog sjældent, at sådanne interventioner bliver givet tilfældigt og under kontrollerede forhold, hvilket kan forklares med etiske og praktiske grunde. (Cattaneo et al., 2017)

Når behandlingen ikke gives tilfældigt kan det føre til en selektion i tildelingen af behandling, og derved opstår der en endogenitetsudfordring, hvis der ikke tages højde for dette i valget af den empiriske strategi. For mit forskningsspørgsmål gør

det sig gældende, at det ikke er muligt, at kontrollere for medførte evner eller IQ. Det ideelle eksperiment for at undersøge ovenstående sammenhæng er, derfor at tage stikprøve i populationen af enægget og tveægget tvillinger, hvor disse skal optages i hver deres optagelseskvote. Tvillingestudier er en måde at korrigere for bias i estimatet. Den basale idé bag denne fixed-effects tankegang er, at nogle af de uobserverbare forskelle vil blive elimineret mellem tvillingepar. (Bingley et al., 2009) Sådanne datasæt er dog meget vanskelige at skaffe, og derfor må der anvendes andre metoder for at løse endogenitetsudfordringen.

Der har i de senere år været anvendt alternative empiriske strategier, som netop forsøger at opnå kausale effekter i fraværet af tilfældig fordelt behandling. En af disse alternative estimationsmetoder er regression diskontinuitet, RD, designet, som har været argumentet som værende en af de mest troværdige ikke-eksperimentelle strategier for at opnå kausale effekter i observationsstudier (Cattaneo et al., 2017).

Den grundlæggende idé i dette design er, at sandsynligheden for at modtage behandling ændres brat ved en kendt grænseværdi, som i mit tilfælde gør sig gældende for optagelse i kvote 1. Hvis det ikke er muligt for individer at sortere sig perfekt omkring denne grænseværdi, antages det at denne diskontinuitets ændring i sandsynligheden kan bruges til at forklare lokale kausale effekter af behandlingen på den ønskede udfaldsvariabel. Individer med en afstandsvariabel, som f.eks. et karaktergennemsnit, der kun marginalt er under grænseværdien, kan benyttes som kontrolgruppe for individer med en afstandsvariabel, som kun er marginalt over. Nedenfor præsenteres det skarpe RD design, som anvendes til at besvare forskningsspørgsmålet.

6.1.1 Det skarpe regression diskontinuitet design

I RD designet vil hvert individ have en afstandsvariabel, og de individer som har en afstandsvariabel over en kendt cut-off vil få tildelt en behandling, mens de som har en afstandsvariabel under denne cut-off ikke vil få tildelt behandling. Disse tre komponenter - afstandsvariablen, cut-off og behandlingen - udgør RD designet. Når alle individer i en undersøgelse modtager den behandling, som de er blevet tildelt på baggrund af deres afstandsvariabel, så anses RD designet som værende skarpt. Modsat gør det sig gældende, at hvis nogle individer ikke modtager behandling, selvom disse har en afstandsvariabel som er over cut-off, eller hvis nogle individer

tildeles behandling, selvom deres afstandsvariabel er under cut-off, så karakteriseres dette RD design for værende *fuzzy*⁴, hvor tildelingen af behandling er mere uklar. (Cattaneo et al., 2017)

Ovenstående kan også formelt udtrykkes ved brug af følgende notation. Det antages, at der er n individer, som er indekseret med $i = 1, 2, \dots, n$. Disse individer har hver en afstandsvariabel, X_i , der i mit set-up er forskellen mellem den enkeltes adgangsgivende karaktersnit og optagelseskvotienten, som er kontinuert. $\bar{x}$ er den kendte cut-off, som er optagelseskvotient til den søgte uddannelse. Individer med $X_i \geq \bar{x}$ vil blive tildelt behandlingstilstanden og således få tildelt en studieplads i kvote 1, mens individer med $X_i \leq \bar{x}$ vil blive tildelt kontroltilstanden, og dermed blive tildelt studieplads i kvote 2. Da data er restrikeret til udelukkende indeholder informationer om optagne studerende i perioden fra 2016-2018, så vil kontroltilstanden i dette empiriske set-up være individer, som er blevet tildelt en studieplads i kvote 2.

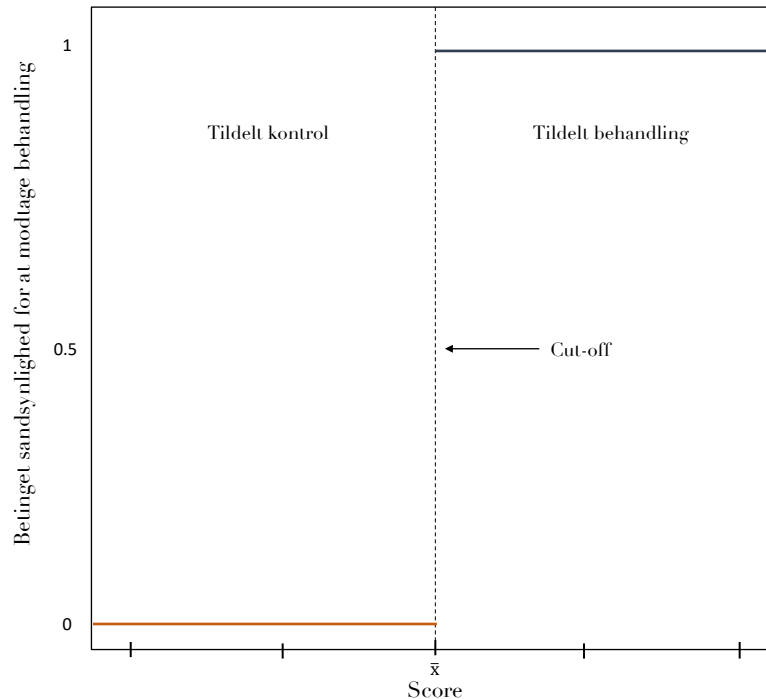
Tildelingen af behandling er betegnet som T_i og er defineret som værende $T_i = \mathbb{1}(X_i \geq \bar{x})$, hvor $\mathbb{1}(\cdot)$ er indikationfunktionen. Denne behandlingsregel medfører, at hvis individernes afstandsvariabel kendes, så vides det også med sikkerhed, om den enkelte er blevet tildelt behandlings- eller kontroltilstanden, hvilket er en af kendetegnene ved RD designet. Sandsynligheden for at modtage behandling som funktion af afstandsvariabel ændres diskontinuert ved cut-off.

Denne behandlingsregel for det skarpe RD design er også grafisk afbildet i figur 6.1. Figuren viser den betingede sandsynlighed, $P(T_i = 1 \mid X_i = x)$, for at få tildelt behandling for en given afstandsvariabel, X_i . I det skarpe RD design er det ikke muligt, at få tildelt behandling, hvis afstandsvariablen ikke er over cut-off, og derfor kan det ses i figuren, at sandsynligheden for at få tildelt behandling ændres fra 0 til 1 ved cut-off (Cattaneo et al., 2017).

I regression diskontinuitet litteraturen antages det, at hver individ vil have to potentielle udfald, $Y_i(1)$ og $Y_i(0)$, som henholdsvis svarer til de observeret udfald hvis individet tildeles behandling eller kontrol. Selvom det antages, at hvert individ har to udfald, så er de kaldt for potentielle idet det kun er muligt at observere et af disse udfald. Hvis individ i tildeles behandling, så vil udfaldet $Y_i(1)$ observeres. Derved

⁴ Se Appendix 11 for yderligere forklaring af det *fuzzy* design

Figur 6.1: Betinget sandsynlighed for at modtage behandling i det skarpe RD design



vil udfaldet $Y_i(0)$ være uobserveret. Omvendt vil det også gøre sig gældende, at hvis individ i tildeles kontroltilstanden vil $Y_i(0)$ observeres, og behandlingstilstanden $Y_i(1)$ vil være uobserveret. Dette gør at det umuligt at opnå behandlingseffekter på individ niveau, da disse er uobserveret (Skovron et al., 2015). Det observeret udfald vil i stedet være givet ved:

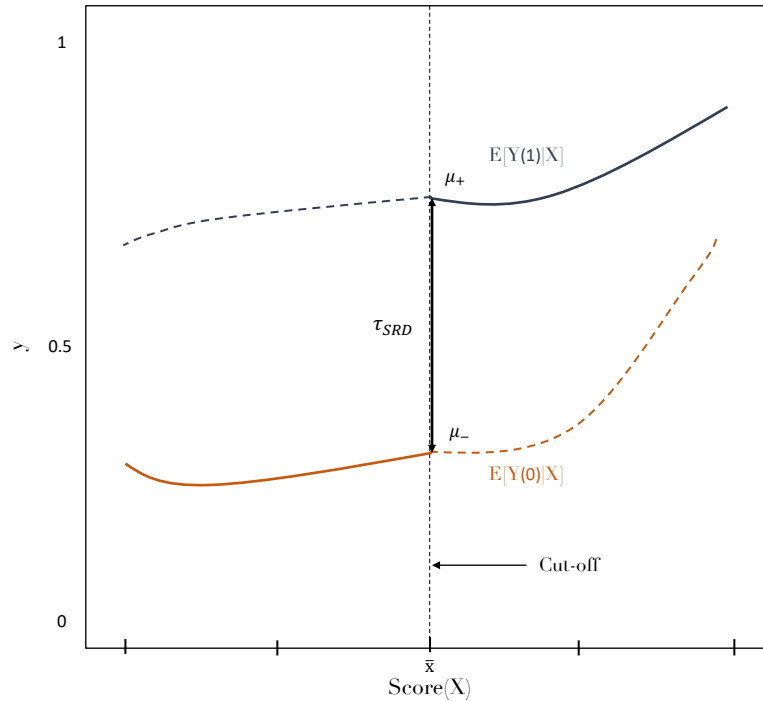
$$Y_i = (1 - T_i) \cdot Y_i(0) + T_i \cdot Y_i(1) = \begin{cases} Y_i(0) & \text{for } X_i < \bar{x} \\ Y_i(1) & \text{for } X_i \geq \bar{x} \end{cases}$$

I figur 6.2 er dette problem vist. Her er de gennemsnitlige potentielle udfald betinget på afstandsvariablen, $E[Y_i(1) | X_i = x]$ og $E[Y_i(0) | X_i = x]$ plottet mod afstandsvariablen. Det kan her ses, at regression funktionen $E[Y_i(1) | X_i]$ er observeret for værdier af afstandsvariablen til højre for cut-off, for når $X_i \geq \bar{x}$ vil det observerede udfald for Y_i være lig det potentielle udfald for behandling for alle $Y_i(1)$. Dette er i figuren vist med en massiv linje. Til venstre for cut-off er alle individer ikke behandlet, og derfor er $E[Y_i(1) | X_i]$ ikke observeret, hvilket i figuren er vist

med en stribet linje. Modsat gør samme problem sig gældende for $E[Y_i(0) | X_i]$, som er observeret for værdier af afstandsvariablen til venstre for cut-off, $X_i < \bar{x}$, men uobserveret for $X_i \geq \bar{x}$. Dette medfører, at de observeret gennemsnitlige udfald er givet ved:

$$E[Y_{ii}] = \begin{cases} E[Y_i(0) | X_i] & \text{for } X_i < \bar{x} \\ E[Y_i(1) | X_i] & \text{for } X_i \geq \bar{x} \end{cases}$$

Figur 6.2: RD behandlingseffekt i det skarpe RD design



Som det kan ses i figur 6.2, så er behandlingseffekten for en given værdi af afstandsvariablen $E[Y_i(1) | X_i = x] - E[Y_i(0) | X_i = x]$ den vertikale afstand mellem de to kurve $\mu_+(\bar{x}) - \mu_-(\bar{x})$. Denne afstand kan dog ikke direkte estimeres, hvilket skyldes at begge kurve aldrig observeres for samme værdi af x . Derfor er det i stedet relevant at kigge på observationer tæt på cut-off med en lille afvigelse på begge sider. Der opstår dog særlig situation ved cut-off. Dette punkt er nemlig det tætteste, som er muligt, at observere begge kurve. Dette er den lokale behandlingseffekt i diskontinuitetspunktet $\bar{x}$ (Cattaneo et al., 2017)

6.1.2 Båndbredde

Båndbredden er meget essentiel i regression diskontinuitet designet, da denne afgør hvilke observationer nær cut-off, som benyttes til at opnå estimationskoefficienten ved brug af en lokale polynomium approksimation. Båndbredden noteres i RD designet som h . En smal båndbredde vil reducere biasen i den lokale polynomium approksimation, men samtidig vil en smal h føre til en stor varians på de estimerede koefficienter, hvilket skyldes at der kun vil være få observationer tilgængelig for estimationen. Omvendt vil en stor båndbredde øge sandsynligheden for bias i den lokale polynomium approksimation, men medføre en lille varians på de estimerede koefficienter. Der vil derfor være valget af båndbredde ofte involvere en bias-variens trade-off. (Skovron et al., 2015)

Da designet er så sensitiv over for valget af båndbredde, anbefales det i eksisterende RD litteratur, at valget af h bør vælges på en datadrevet måde. En tilgang til bruge denne datadrevet fremgangsmåde er, at vælge en båndbredde som minimere den asymptotiske mean squared error, MSE, af RD punktestimatoren. Denne fremgangsmåde vælger h til at optimere bias-variens trade-off'et, således at der laves en afvejning mellem de to. (Lee & Lemieux, 2010)

I afsnit 8 vil sensibiliteten af estimererne i forhold til valget af båndbredde blive testet for at undersøge, hvor robust af de fundne resultater er, når båndbredden ændres.

6.1.3 Antagelser

RD designet anses for at være stærkt design for et opnå kausale effekter. Designets antagelser er desuden tiltalende idet, at disse er relativt milde sammenlignet med andre ikke-eksperimentielle tilgange. Nogle af disse antagelser er mulige at teste, mens andre ikke er mulige. Dog er det en nødvendighed, at antagelserne er opfyldt for at kunne fortolke resultaterne for kausale. (Lee & Lemieux, 2010) Antagelser for bag designet vil her blive introduceret:

1. Der skal være en diskontinuitet i sandsynligheden for tildelingen af behandling i grænseværdien

En fundamental antagelse i RD designet er, at der skal være en diskontinuert ændring i sandsynligheden for at modtage behandling ved cut-off. Denne an-

tagelse er mulig at teste ved et grafisk plot. Det er per definition ikke muligt, at blive optaget i kvote 1, hvis den studerende ikke har et karaktergennemsnit fra en ungdomsuddannelse, som er lig eller højere end grænsekotienten, hvorfor at det ikke er overraskende, at der er en klar sammenhæng mellem de to variable.

2. Individerne må ikke kunne manipulere med deres afstandsvariabel

Hvis individer har upræcis kontrol over afstandsvariablen, X , anses behandling for værende "så god som" tilfældig fordelt omkring cut-off. Studerende som eksempelvis kan manipulere med deres afstandsvariabel X gennem deres indsats og derved opnår en afstandsvariabel større end $\bar{x}$, som medfører, at de modtager en behandling, antages for at være forskellige fra de studerende, som vælger ikke at manipulere med deres afstandsvariabel gennem en indsats, så de ender lige under $\bar{x}$. (Lee & Lemieux et al., 2010)

Det er således værd at være opmærksom på, at selvom der er en diskontinuitet i tildelingen af behandling, så er dette ikke tilstrækkeligt nok til at kunne argumentere for et validt RD design. I dette speciale anvendes det institutionelle set-up, hvor tildelingen af studiepladser foregår gennem et centraliseret system, som beskrevet i afsnit 4. Grænsekotienterne til de danske videregående uddannelser, bliver som tidligere beskrevet, dannet hvert år, hvorfor de er ikke præcist kendes af ansøgerne på ansøgningstidpunktet. Dette skaber en tilfældig variation om, hvilken kvote de studerende optages i.

3. Individer tæt på cut-off antages for ens, i gennemsnit, på observerede og uobserverede karakteristika

I RD designet betyder dette, at fordelingen af de observerede og ikke-observerede karakteristika skal være kontinuerlig omkring cut-off. Selvom det kan være muligt at kontrollere for observerede kovariater, skal ligheden mellem ikke-observerede egenskaber antages at være ens. Dette betragtes som en sandsynlig antagelse for individer meget tæt på cut-off punktet. (Lee & Lemieux et al., 2010).

Denne antagelse kan testes ved inklusion af kovariater i estimationen. I RD designet antages det, at inklusion af disse kovariater ikke vil ændre de estimerede koefficienter markant. Denne antagelse vil senere blive testet i afsnit

6.2 Økonometrisk specifikation

Den empiriske analyse vil anvende det beskrevne regression diskontinuitet design. RD designet estimeres ved brug af OLS I dette speciale vil afstandsvariablen være distancen mellem den enkeltes karaktergennemsnit fra en ungdomsuddannelse og grænsekvtienten, som kan udtrykkes ved følgende:

$$x_i = KG_i - GK_i \quad (6.1)$$

Her er KG_i det adgangsgivende karaktergennemsnittet fra en ungdomsuddannelse for individ i , og GK_i er grænsekvtienten for individ i 's optagne uddannelse i et givent år. $x_i = 0$ vil anvendes som cut-off, hvilket betyder, at et individ nøjagtigt har det krævet snit for at blive optaget. Hvis $x_i > 0$ vil dette betyde, at den enkelte har et adgangsgivende karaktergennemsnittet, som er højere end den krævede kvotient, og derfor vil denne have opnået optagelse i kvote 1. Modsat gør det sig gældende hvis $x_i < 0$, så vil individet have et adgangsgivende karaktergennemsnittet som er lavere end den krævede kvotient, og derfor har denne opnået optagelse i kvote 2. De forskellige grænsekvtienter er normaliseret, således at der kun finde én cut-off værdi ved $x_0 \equiv x_i = 0$. Designet er skarpt, som nævnt i afsnit 6.1.1, da optagelsesstatussen ændres deterministisk som funktion af afstandsvariablen. Dette skyldes, at der i data udelukkende indeholder informationer om optagne studerende og ikke ansøgere.⁵ Behandlingsgruppen består derfor af individer, som er optaget på en given uddannelse, hvor deres karaktergennemsnit udelukkende er anvendt, mens kontrolgruppen af individer består af optagne, hvor deres karaktergennemsnit ikke opfylder den given grænsekvtient for uddannelsen.

Derved kan sandsynligheden for frafald for den enkelte optagne studerende, i således udtrykkes ved følgende ligning:

$$y_i = \alpha_1 + \delta T_i + f(x_i) + \epsilon_i \quad (6.2)$$

hvor y_i er frafaldssandsynligheden for den enkelte studerende, $f(\cdot)$ er en fleksibel

⁵ Se mere i appendix 11

funktion, som er et polynomium af graden p , som fanger effekten afstandsvariablen, hvorfor denne funktion giver en god approksimation af de underliggende variable, og dermed $E[y_i|x_i]$. (Angrist & Pischke, 2009) ϵ_i er fejleddet. Dette er det simple RD design, hvor effekten af optagelsesknoten på frafaldsandsynligheden undersøges. Denne model vil være udgangspunktet for dette speciale. Dog udvides modellen også, således den vil indeholde kontroller på baggrundskaraktistika. Derved vil den udvidede model være givet ved følgende:

$$y_i = \alpha_1 + \beta K_i + \delta T_i + f(x_i) + \epsilon_i \quad (6.3)$$

Her er K_i er en vektor af kontrolvariable såsom alder, køn, forældres uddannelsesniveau og forældres indkomst, målt i log. For at teste hvor robust de empiriske resultater er, så inkluderes der også en kontrol for de studerendes selvrapporterede afklarethed, som beskrevet i afsnit 5.1.5. Dette medfører, at denne model vil være givet ved:

$$y_i = \alpha_1 + \beta K_i + \delta T_i + f(x_i) + \lambda a_i + \epsilon_i \quad (6.4)$$

Hvor λa_i angiver de studerendes selvrapporterede niveau af afklarethed. Der er hermed præsenteret 3 forskellige økonometriske specifikationer af den empiriske model. Disse vil afsnit 7 blive estimeret. Det er dog muligt, at der vil være andre faktorer, som kan have betydning for udfaldsvariablen. Derfor vil der i det kommende afsnit blive præsenteret mulige udeladte variable, som kan have betydning for estimerterne.

6.3 Udeladte variable

Det er plausibelt, at nogle uddannelser adskiller sig på visse parametre, som eksempelvis uddannelser med et højere adgangskrav, hvilket kan være med til at påvirke de studerendes vedholdenhed. Som tidligere nævnt i afsnit 3 spiller faktorer som deltagelse i studiegrupper, social integration og faglig integration også en stor rolle for, hvem der falder fra i løbet af det første studieår. Disse faktorer er ofte påvirket af, hvilken uddannelse og institution der er tale om. Derudover er de uddannelsesspecifikke karakteristika svære, hvis ikke umulige, at medtage i en empirisk analyse, hvorfor de vil som oftest blive set som uobserverbare faktorer.

Såfremt at der er uddannelsesspecifikke karakteristika, som påvirker sandsynligheden for frafald vil estimatet af optagelseskvotes effekt på frafald være biased, da en del af effekten af uddannelsesspecifikke fejlagtigt bliver tilskrevet optagelseskvoten i modellen. Med andre ord vil udeladelse af blandt andet uddannelsesspecifikke karakteristika betyde, at jeg vil over- eller undervurderer effekten af optagelseskvoten. Det vil altså sige, at udeladelsen af visse variable har konsekvenser for gyldigheden og troværdigheden af mine modeller. (Angrist & Pische, 2009)

Hvis der ikke bliver kontrolleret for en udeladt variabel, er den indeholdt i fejleddet, der hermed kommer dette til at bestå af både den udeladte variabel og det strukturelle fejldet. En løsning på dette problem kan være fixed effects. Derfor præsenteres denne metode, som netop er en metode, som tager højde for uobserverbare faktorer i modellen. (Wooldridge, 2009) Nedenfor præsenteres fixed effects, FE. Denne metode tager netop højde for uobserverbare faktorer i modellen.

6.3.1 Fixed Effects

For at tage hensyn til mulige udeladte variable gennemføres også estimationer, hvor der anvendes Fixed Effects, FE, som netop tager højde for uobserverbare faktorer i modellen (Wooldridge, 2009). Ved brug af fixed effects anvendes fordelene ved at datasættet er longitudinalt. Det overordnede mål med FE er, at tillade for at uobserverede effekter er korreleret med en eller flere af de forklarende variable. Paneldata tillader os hermed, at kontrollere for, at dette kan være tilfældet (Wooldridge, 2009). Det vil dermed være muligt, at se bort fra variationen mellem uddannelserne for derved at tage højde for, at bestemte typer af studerende er tilbøjelige til at søge ind på bestemte uddannelser. På den måde sammenlignes kun studerende inden for samme specifikke uddannelsesudbud. Med udgangspunkt i ligning 6.2 opstilles FE-modellen nedenfor:

$$y_{it} = \alpha_1 + \delta T_{it} + f(a_{it}) + \epsilon_{it} \quad (6.5)$$

Hvor i angiver hvert individ og t angiver uddannelsesudbuddet. Fejleddet ϵ_{it} er opdelt i to:

$$\epsilon_{it} = k_i + u_{it} \quad (6.6)$$

Hvor k_i er de konstante uobserverbare faktorer, og u_{it} er det strukturelle fejled. (Woolridge, 2009) FE-modellen anvendes til, at undersøge forholdet mellem den forklarende variabel og den afhængige variabel inden for hvert uddannelsesudbud. Derved opnås der *within* effekter, hvor hvert uddannelsesudbud har sit eget sæt af uddannelsesspecifikke karakteristika, som enten påvirker eller ikke påvirker den afhængige variabel. For at fjerne dette bias trækkes gennemsnittet fra uddannelsesudbuddet fra, og dermed eliminerer jeg den uddannelsesspecifikke konstante effekt, k_i :

$$\begin{aligned}(y_{it} - \bar{y}_i) &= \alpha_1 - \alpha_1 + \delta(T_{it} - \bar{T}_{it}) + f(a_{it} - \bar{a}_{it}) + (k_i - \bar{k}_i) + (u_{it} - \bar{u}_i) \\ \Leftrightarrow \dot{Y}_{it} &= \delta \dot{T}_{it} + f(\dot{a}_{it}) + \dot{u}_{it}\end{aligned}\tag{6.7}$$

For at opnå en unbiased FE estimation kræver det, at der er flere antagelser, som skal være opfyldt. Det gælder blandt andet, at der skal være en tilfældig udvælgelse fra cross section (FE.2), hvor hver forklarende variabel ændrer sig over uddannelsesbuddene, som minimum for nogle i , og desuden må der ikke være perfekt lineære forhold mellem de forklarende variable (FE.3) (Woolridge, 2009)

Ligeledes vil modellerne i ligning 6.3 og 6.4 blive estimeret i afsnit 7 som fixed effects, således at studerende på samme specifikke uddannelsesudbud sammenlignes.

7 Resultater

I dette afsnit præsenteres hovedresultaterne af den empiriske analyse. Inden disse præsenteres, vil den interne validitet af RD designet gennemgås. Herefter præsenteres resultaterne af hvordan optagelseskvoten påvirker den tidlige frafaldssandsynlighed. Efterfølgende undersøges potentielle mekanismer bag effekterne mere detaljeret ved at fokusere på sub-populationer. Afslutningsvis indeholder afsnittet et overblik over hovedresultaterne.

7.1 Intern validitet af det empiriske design

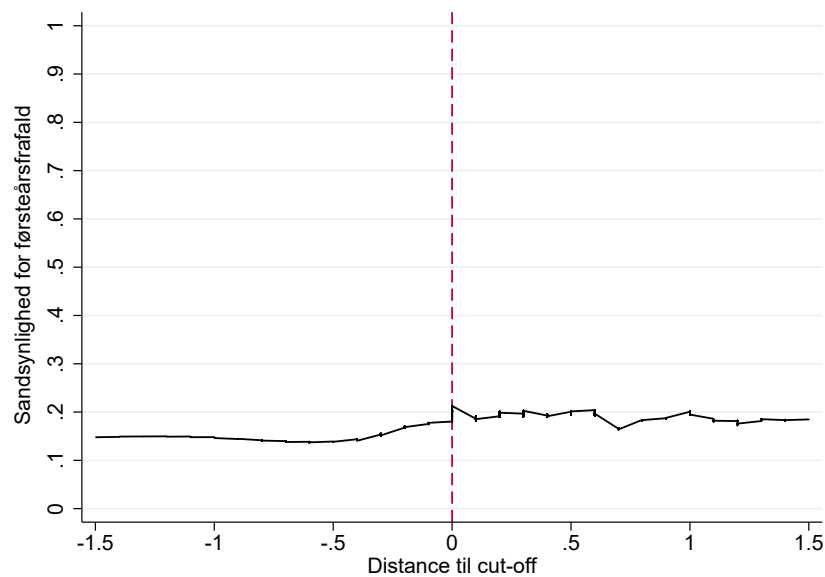
En væsentlig betingelse for det empiriske design er, at optagelsessystemet til universitetsuddannelserne skaber et hop i sandsynligheden for at falde fra som funktion af afstandsvariablen. For at RD designet kan anvendes, er det nødvendigt, at der er en observerbar tildelingsvariabel. Denne tildelingsvariabel er karakteriseret ved, at tildelingen af behandling afhænger af den enkeltes værdi for tildelingsvariablen. I mit speciale er tildelingsvariablen baseret på afstanden mellem de studerendes karaktergennemsnit fra en ungdomsuddannelse og grænsekvotienten på en given uddannelse for et givent år, som også beskrevet i ligning 6.1.

En grafisk inspektion af data præsenteres her i figur 7.1. Denne inspektion foretages for at undersøge, om der findes en diskontinuitet ved cut-off, hvilket retfærdiggøre brugen af RD designet, som også nævnt i afsnit 6.1.3. Figur 7.1 viser et simpelt og råt plot af sammenhængen mellem frafald og afstanden til cut-off.

Distancen til cut-off er afbildet på x-aksen i figur 7.1, denne er dannet som angivet i ligning 6.2, hvor viser forskellen mellem det adgangsgivende karaktergennemsnit som den studerende har opnået fra en ungdomsuddannelse og optagelseskvotienten for den søgte uddannelse. Alle optagelseskvotienter er blevet poollet og normaliseret. En distance på 0 på x-aksen betyder derfor, at den studerende har præcis et adgangsgivende karaktersnit svarende til adgangskvotienten på den pågældende uddannelse. Dette betyder også, at de studerende, som er placeret til venstre og har en negativ afstand til cut-off vil være optaget i kvote 2 på den pågældende uddannelse, mens alle studerende, som er placeret til højre og har en positiv afstand til cut-off vil være optaget i kvote 1 på den pågældende uddannelse.

Figur 7.1 viser, at frafaldssandsynligheden for de studerende, som har et karaktergennemsnit der er højere end adgangskvotienten, og derfor er optaget i kvote

**Figur 7.1: Sammenhængen mellem optagelseskvote og
frafaldssandsynligheden for universitetsbacheloror, 2016-2018**



Note: Figuren viser en simpel sammenhæng mellem at frafalde påbegyndt uddannelse inden for et studieår for studerende optaget i kvote 2 (venstre) og kvote 1 (højre). Figuren viser kun individerne med et karaktersnit 1,5 karakter fra cut-off. Der er i figuren anvendt den lokal vægtet regression scatterplot

Kilde: Egne beregninger baseret på registerdata fra Danmarks Statistik

1, har en højere sandsynlighed for at falde fra studiet i løbet af det første studieår sammenlignet med de studerende, som har et karaktersnit der er lavere end adgangskvotienten, og derfor er optaget i kvote 2. Denne simple figur bekræfter dermed den formodede sammenhæng og hypotese. Mere væsentlig ser det også ud til, at der forekommer en diskontinuitet i frafaldssandsynligheden ved cut-off. Dette ses ved, at der forekommer et væsentlig hop ved nul på x-aksen i frafaldssandsynligheden.

Dette tegn på en diskontinuitet styrker validiteten af det empiriske design, hvorfor der i de efterfølgende afsnit udføres forskellige RD regressioner.

7.2 Optagelseskvotes betydning på frafaldssandsynligheden

I dette afsnit vil regressionsresultaterne af effekten af optagelseskvoten på frafaldssandsynligheden præsenteres. Første gives der en visuel repræsentation af resultaterne fra ligning 6.2, som er det simple RD design, samt ligning 6.3.1, hvor benyttes fixed effects på uddannelsesudbuddet.

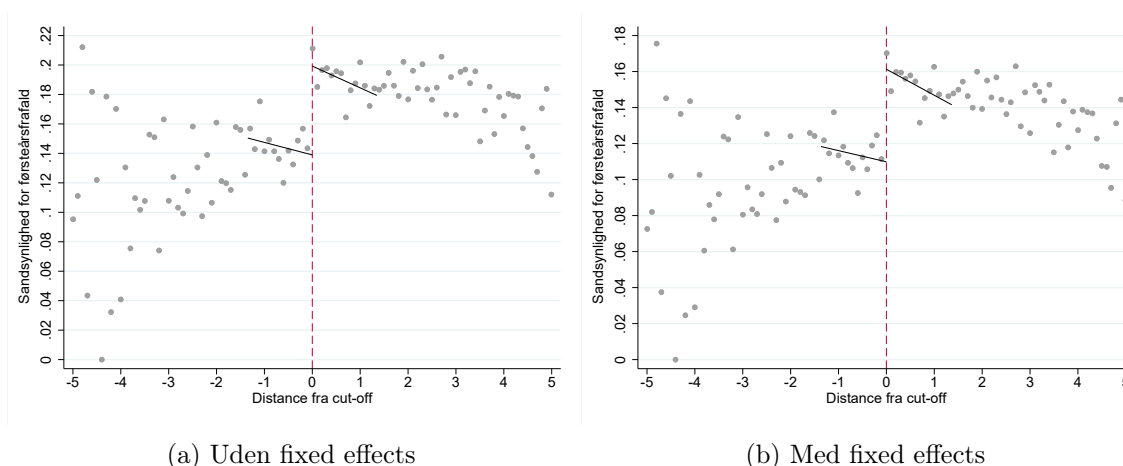
I figur 7.2a, som er den simple RD model uden fixed effects, kan det ses, at sker et tydeligt hop i frafaldet for de to grupper. Figuren viser, at studerende optaget i kvote 1 vil opleve et væsentligt højere frafald sammenlignet med studerende optaget i kvote 2. Her kan det ses, at den lokale gennemsnitlige effekt på frafaldssandsynligheden af at være optaget i kvote 1 på en universitetsbachelor er på omtrent 5,5 pct. point højere sammenlignet med studerende, som er optaget i kvote 2. Den stiblet vertikale linje markerer cut-off ved nul. I figuren er anvendt den MSE-optimale båndbredde, som beskrevet i afsnit 6.1.2 formår denne at minimere afvejningen mellem præcise estimer og bias.

Resultatet kan ligedes ses i tabel 7.2 i kolonne (1) . Her kan det ses, at der er anvendt MSE optimal båndbredde på 1,347 for estimatet. Denne forskel i frafaldssandsynligheden er af en væsentlig størrelse, og dette resultat bekræfter dermed hypotesen om, at studerende optaget i kvote 2 vil falde mindre fra studiet i løbet af det første studieår.

Figur 7.2b er RD modellen, hvor der er indført fixed effects på uddannelsesudbuddet, vist. Denne tager højde for variationen inden for samme uddannelsesudbud. I denne figur kan det ses, at forskellen i frafaldssandsynligheden vedbliver, og der dermed er en større frafaldssandsynlighed for studerende, som er optaget i kvote 1

sammenlignet med kvote 2. Det kan af tabel 7.2 kolonne (2) aflæses, at den estimerede forskel er på 4,8 pct. point. Hvilket betyder, at de studerende, som er optaget i kvote 1 har en 4,8 pct. point større sandsynlighed for at falde fra i løbet af det første studieår.

Figur 7.2: Optagelseskvotes effekt på førsteårs frafaldssandsynligheden for universitetsbacheloror, 2016-2018



Note: Figuren viser en simpel sammenhæng mellem at frafalde påbegyndt uddannelse inden for et studieår for studerende optaget i kvote 2 (venstre) og kvote 1 (højre).

Kilde: Egne beregninger baseret på registerdata fra Danmarks Statistik

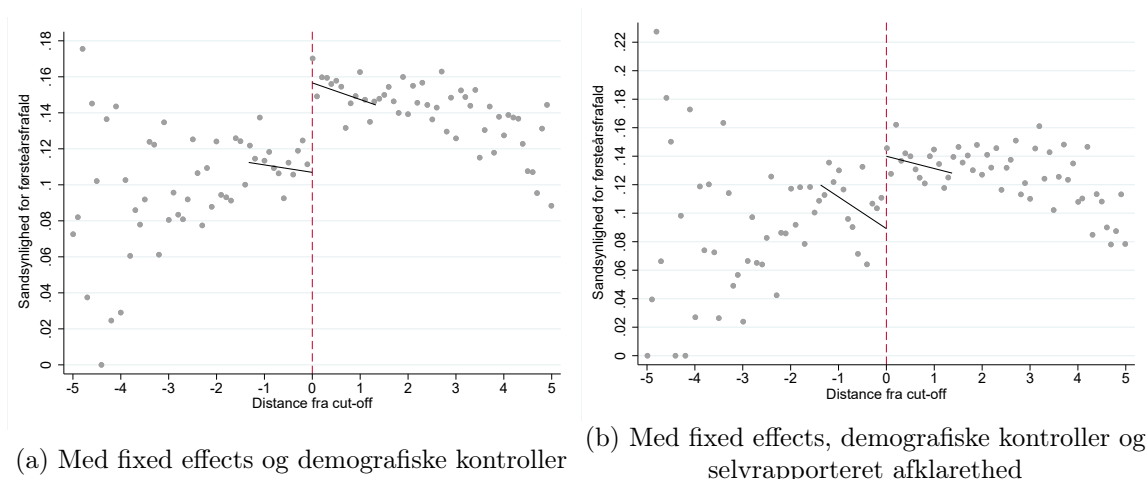
Der er en forskel på 0,7 mellem de to estimater for førsteårsfrafaldet mellem to modeller. Det må dermed påstås, at der er nogle uddannelsesspecifikke forhold som gør sig gældende, hvorfor vil den mest retvisende model og resultat må være det, som opnås med fixed effects RD modellen. Derfor vil denne model med fixed effects anvendes, når modellen nu udvides med kovariater.

7.2.1 Inklusion af kovariater

I dette afsnit præsenteres estimaterne fra ligning 6.3 og 6.4, hvor der ligeledes vil være fixed effects på uddannelsesudbuddet. Først inkluderes demografiske kovariater, som beskrevet i afsnit 5, og derefter tilføjes der desuden også et mål for selvrapporteret afklarethed. I RD designet bør inklusionen af kovariater ikke påvirke estimaterne betydeligt, da kovariaterne bør være tilfældige omkring diskontinuiteten, som beskrevet i afsnit 6. Det er en af de grundlæggende antagelser bag designet, at behandlingen gives tilfældigt, og derfor bør der ikke være store forskelle i observerbar karakteristika. Inklusionen sker dog alligevel for at styrke troværdigheden

for de opnåede resultater.

Figur 7.3: Optagelseskvotes effekt på førsteårs frafaldssandsynligheden for universitetsbachelorer, 2016-2018



Note: Figuren viser en simpel sammenhæng mellem at frafalde påbegyndt uddannelse inden for et studieår for studerende optaget i kvote 2 (venstre) og kvote 1 (højre).

Kilde: Egne beregninger baseret på registerdata fra Danmarks Statistik

Figur 7.3a viser RD modellen hvor der er indført fixed effects på uddannelsesudbudet og demografiske kovariater. I denne figur kan det ses, at forskellen i frafaldssandsynligheden vedbliver, og den estimerede forskel er på 4,5 pct. point. Hvilket betyder, at de studerende, som er optaget i kvote 1 har en 4,5 pct. point større sandsynlighed for at falde fra i løbet af det første studieår. Tilføjelsen af kovariater ændrer kun ganske lidt på den estimerede effekt, som kan ses i figur 7.2b. Denne lille ændring i estimerterne tyder dermed på, at de studerende nær cut-off er ens på tværs af disse kovariater.

Figur 7.3b viser RD modellen hvor der er indført fixed effects på uddannelsesudbudet, demografiske kovariater samt det selvrapporterede mål for afklarethed. Figuren viser, at der fortsat er en forskel i frafaldssandsynligheden. Den estimerede forskel er på 4,3 pct. point. Hvilket betyder, at de studerende, som er optaget i kvote 1 har en 4,3 pct. point større sandsynlighed for at falde fra i løbet af det første studieår. Tilføjelsen af de demografiske kovariater samt detselvrappede niveau af afklarethed har dermed kun haft en ganske lille betydning for estimatet for frafaldssandsynligheden.

Resultaterne for de fire modeller kan ses i tabel .5. Det kan af denne aflæses, at den MSE optimale båndbredde afhænger af modellen. Dette skyldes, at antallet af observationer varierer på tværs af modellerne.

**Tabel 7.1: Effekten af optagelseskvoten på førsteårs
frafaldssandsynlighed for universitetsbachelorere, 2016-2018**

	Simpelt RD	Fixed effects RD	Fixed effects RD m/ demografiske kontrol	Fixed effects RD m/ demografiske kontrol og afklarethed
	(1)	(2)	(3)	(4)
RD koeficient	0,055*** (0,014)	0,046*** (0,011)}	0,043*** (0,011)	0,042*** (0,015)
Båndbredde, h	1,347	1,366	1,344	1,369
Demografiske kontroller	Nej	Nej	Ja	Ja
Selvrapporteret afklarethed	Nej	Nej	Nej	Ja
Fixed effects	Nej	Ja	Ja	Ja
Antal effektive observationer	27.124	27.124	27.124	13.367
Til venstre for cut-off	3.773	3.773	3.773	1.839
Til højre for cut-off	23.351	23.351	23.351	11.528

Note: Robuste standardfejl er angivet i parentes. * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$. Båndbredde er valgt ved brug af MSE.

Kilde: Egne beregninger baseret på registerdata fra Danmarks Statistik

På tværs af disse 4 modeller vil studerende, som er optaget i kvote 2 have en væsentlig lavere frafaldssandsynlighed sammenlignet med studerende optaget i kvote 1. Resultaterne giver ikke en direkte forklaring på, hvilke årsager som gør, at de studerende i kvote 2 er bedre polstret til at undgå at falde fra sin uddannelse. Disse resultater kan tyde på, at optagelsen i kvote 2 giver de studerende en form for 'boost' eller en form for øget motivation for at gennemføre uddannelsen, hvorfor noget tyder på, at kvote 2 kan være et middel for at sikre en højere gennemførsel.

I modellen med fixed effects, demografiske kontroller og selvrapporterede afklarethed vil båndbredden være størst. Dette skyldes, at der i denne model er færrest observationer, så der vil derfor være behov for flere observationer for at give de mest præcise estimater. Som nævnt tidligere i 5.1.5, så er det omtrent halvdelen af de optagne, som har valgt af svare på dette spørgeskemaet, hvilket også kan ses i antallet

af observationer.

For at undersøge om, der er nogle bestemte typer af studerende, som i særlig grad driver disse effekter, så undersøges de heterogene effekter nu. Som baseline for disse undersøgelser vil den simple RD model med fixed effects blive anvendt.

7.3 Heterogenitet i treatment-effekter

Resultaterne viser, at studerende, som optages i kvote 2 er mere vedholdende end de studerende, som optages i kvote 1. I dette afsnit undersøges det, hvordan effekten af optagelseskvoten varierer for subpopulationer. Først undersøges effekterne af optagelseskvoten på de studerendes frafaldssandsynlighed, når der betingelse på udvalgte baggrundskaraktéristika, dernæst belyses de heterogene effekter på uddannelseskaraktéristika og slutteligt vil de heterogene effekter af kriterierne for kvote 2 optagelse undersøges.

7.3.1 Effekter for subpopulationer defineret ved baggrundskaraktéristika

Dette afsnit belyser, hvordan effekterne af optagelseskvoten på frafaldssandsynligheden varierer, når der betingelse på hhv. køn, alder og forældres uddannelsesniveau.

I tabel 7.2 kolonne A.1 kan undergrupperne opdelt i køn ses. Her kan det bemærkes, at for både kvinder og mænd gør det sig gældende, at optagelse i kvote 1 er forbundet med en øget sandsynligheden for at falde fra i løbet af det første studieår sammenlignet med kvote 2. Dette er i overensstemmelse med tidligere fund. Dog er det bemærkelsesværdigt, at frafaldssandsynlighed er over tre gange så stor for mænd sammenlignet med kvinder. Dette indikerer, at mænd er mere tilbøjelig til opgive deres studie, når de bliver optaget i kvote 1 sammenlignet med de mænd, som bliver optaget i kvote 2. Mænd har på tværs af alle uddannelsesforløb har et højere frafald sammenlignet med kvinder, hvorfor det ikke er så overraskende, at estimatet er højere for mænd end kvinder. Dog er det interessant, at effekten er så stor for mænd. Denne effekt er også statistisk signifikant, hvilket indikerer, at mænd i højere grad oplever et boost eller bliver mere motiveret af at blive optaget i kvote 2 sammenlignet med de mænd, som bliver optaget i kvote 1.

Derudover kan det af kolonne B aflæses, at alderen har en betydning for frafaldssandsynligheden i kvote 1 og kvote 2. Det kan af linje B.1 aflæses, at der ikke findes

**Tabel 7.2: Heterogene effekter af optagelsesknoten på studerendes
frafaldssandsynlighed for subpopulationer defineret ved
baggrundskaraktetika**

	RD koefficient	Båndbredde	Antal effektive observationer
Baseline estimeret fra tabel 7.1 i kolonne (2)	0,048*** (0,011)	1,366	27.124
A. Køn			
A.1 Kvinde	0,024* (0,014)	1,680	17.386
A.2 Mand	0,075*** (0,015)	1,592	13.226
B. Alder			
B.1 Under 20 år	0,008 (0,038)	1,092	3.365
B.2 20-22 år	0,051*** (0,115)	1,197	15.823
B.3 23-25 år	0,044* (0,024)	1,992	4.509
B.4 Over 25 år	0,059* (0,033)	1,581	1.950
C. Forældres uddannelse			
C.1 Mindst én forældre har en lang videregående uddannelse	0,080*** (0,019)	0,983	7.442
C.2 Forældre har ikke en lang videregående uddannelse	0,031** (0,014)	1,424	18.106

Note: RD estimeret med fixed effects. Robuste standardfejl er angivet i parentes. * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$

Kilde: Egne beregninger baseret på registerdata fra Danmarks Statistik

nogle signifikant forskel i frafaldssandsynligheden for studerende optaget i kvote 1 og kvote 2 for de, som er under 20 år. Dette kan også forklares med, at dette er en lille gruppe af de studerende. Dog er effekterne af optagelseskvoten på frafaldssandsynligheden signifikant for de resterende aldersgrupper. Særligt stor er forskellen for aldersgruppen over 25 år. Her vil de over 25 år, som er optaget i kvote 1 have en frafaldssandsynlighed, som er 5,9 pct. point større end de over 25 år optaget i kvote 2. Mens studerende optaget i kvote 1 i aldersgrupperne 20-22 år og 23-25 år vil have en frafaldssandsynlighed som er hhv. 5,1 pct. point og 4,4 pct. point større end studerende optaget i kvote 2 i samme aldersgruppe.

Derudover kan det af tabel 7.2 aflæses, at forældres uddannelsesbaggrund af betydning for frafaldssandsynligheden. Studerende, som er optaget i kvote 1, som har mindst én forældre med en lang videregående uddannelse, vil have 8 pct. point større frafaldssandsynlighed sammenlignet med en studerende optaget i kvote 2 med samme en forældre med samme uddannelsesniveau. Denne forskel er væsentlig over baseline estimatet på 4.8 pct. point. Studerende med mindst én forældre med en lang videregående uddannelse vil opleve et stort 'boost' af optagelsen i kvote 2. Denne effekt er ikke nær så stærk for studerende, hvis forældre ikke har en lang videregående uddannelse. Disse studerende, som er optaget i kvote 1, vil have 3,1 pct. point større frafaldssandsynlighed sammenlignet med studerende, som er optaget i kvote 1, hvor forældrene ikke har en lang videregående uddannelse. Dette resultat er under baseline estimatet.

7.3.2 Subpopulation effekter defineret ved uddannelseskaraktistika

Dette afsnit belyser, hvordan effekterne af optagelseskvoten på frafaldssandsynligheden varierer, når der betinges på hhv. de studerendes optagelsesprioritet og den optagne uddannelses grænsekotient.

Det kan af tabel 7.3 aflæses, at studerende optaget i kvote 2 på deres 2. prioritet falder mindre fra sammenlignet med studerende optaget i kvote 1 også på deres 2. prioritet. Derudover viser tabellen i afsnit B, at studerende som er optaget i kvote 2 på en uddannelse med en grænsekotient over 7 oplever et væsentlig lavere frafaldssandsynlighed sammenlignet med studerende, som er optaget på de samme uddannelser bare i kvote 1. Dette resultater tyder på, at optagelse på en uddannelse med en høj grænsekotient i kvote 2 giver et væsentlig 'boost' til de studerendes

**Tabel 7.3: Heterogene effekter af optagelsesknoten på studerendes
frafaldssandsynlighed for subpopulationer defineret ved
uddannelseskarakteristika**

	RD koefficient	Båndbredde	Antal effektive observationer
Baseline estimater fra tabel 7.1 kolonne (2)	0,048*** (0,011)	1,366	27.124
<i>A. Optagelsesprioritet</i>			
<i>A.1</i> 1. prioritet	0,042*** (0,014)	1,218	18.233
<i>A.2</i> 2. prioritet	0,052*** (0,022)	1,602	6.288
<i>B. Grænsekotient</i>			
<i>B.1</i> Under 4	0,309*** (0,110)	0,373	63
<i>B.2</i> 4-6,99	0,050*** (0,028)	1,176	4.099
<i>B.3</i> 7-9,99	0,062* (0,016)	1,321	15.379
<i>B.4</i> 10 eller derover	0,060* (0,020)	1,341	5.310

Note: RD estimaterne er esitmeret med fixed effects. Robuste standardfejl er angivet i parentes. * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$

Kilde: Egne beregninger baseret på registerdata fra Danmarks Statistik

vedholden. Bemærkelsesværdigt er det også, at dette 'boost' også er at finde for studerende optaget i kvote 2 for uddannelser med en grænsekotient mindre end 4. Dog bør det holdes for øje, at antallet af observationer for denne gruppe er meget lav, hvorfor resultatet bør fortolkes med varsomhed.

7.3.3 Subpopulation effekter defineret ved kvote 2 kriterier

I tabel 7.4 er de heterogene effekter for de tre kriterier, som er blevet vurderet for optagelse i kvote 2, vist. Her kan det ses, at særligt de studerende, som er optaget i kvote 2, hvor det har været en passiv med en motiveret ansøgning har en langt større sandsynlighed for at blive på studiet sammenlignet med de studerende, som er optaget på samme studie i kvote 1. Det er dog værd, at have in mente, at dette kriterie er et af de mest anvendte. Det ses også umiddelbart, at antallet af observationer er størst her.

Tabel 7.4: Heterogene effekter af optagelsesknoten på studerendes frafaldssandsynlighed for subpopulationer defineret ved uddannelseskarakteristika

	RD koefficient	Båndbredde	Antal effektive observationer
Baseline estimeret fra tabel 7.1 kolonne (2)	0,048*** (0,011)	1,366	27.124
A. Kvote 2 kriterier			
A.1 Passiv uden motiveret ansøgning	0,032 (0,020)	1,273	9.187
A.2 Passiv med motiveret ansøgning	0,080*** (0,018)	0,018	5.112
A.3 Fremmøde	0,035* (0,021)	2,172	4.990

Note: RD estimeret med fixed effects. Robuste standardfejl er angivet i parentes. * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$

Kilde: Egne beregninger baseret på registerdata fra Danmarks Statistik

7.3.4 Opsummering

Resultaterne af analyserne bekræfter den formodede sammenhæng og hypotese om, at der findes en klar effekt af optagelseskvote på frafaldssandsynligheden. Studerende, som optages i kvote 2, har en væsentlig lavere og signifikant frafaldssand-

synlighed. Denne forskel vedbliver selvom, at der inkluderes kovariater. Resultaterne siger ikke noget om, hvilke årsager som ligger bag den lavere frafaldssandsynlighed for studerende optaget i kvote 2. Men en mulig årsag kan være, at optagelsesredskaberne i kvote 2 er bedre til at matche ansøger og uddannelser, samt selve optagelsesprocessen giver den studerende et form for 'boost' som øger motivationen. Resultaterne i kapitlet indikerer, at det kan være et nyttigt politisk værktøj at mindske frafaldet ved at øge optaget i kvote 2. For at verificere de opnåede resultater er det nødvendigt at gennemføre flere robusthedscheck af designet. Det følgende afsnit vil gennemgå relevant robusthedscheck. Til disse vil den primære model fixed effects RD model anvendes.

8 Robusthedscheck

I dette afsnit undersøges robustheden af de fundne resultater i afsnit 7. I RD designet er den interne validitet baseret på en lokal kontinuitets antagelse, som tilsiger at både uobserverbare og observerbare faktorer er ens på begge sider af cut-off, således vil det kun være tildelingen af behandling, som skaber denne diskontinuitet. Denne antagelse kan ikke direkte testes, men i stedet er det muligt, at anvende en række andre test, som kan underbygge plausibiliteten af denne antagelse. Dette vil gøres ved brug af tre forskellige empiriske test, som vil kunne besvare, hvor robuste de fundne resultater er.

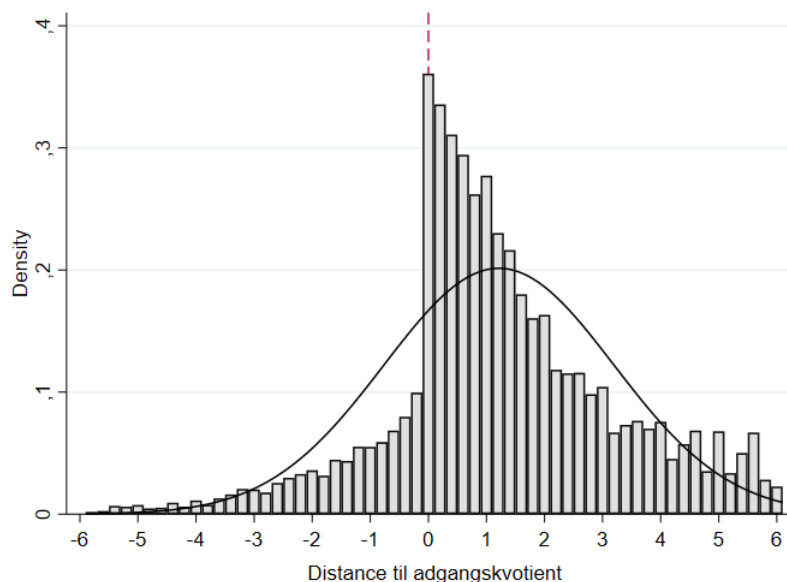
8.1 Fordelingen af afstandsvariablen

En anden form for validitetstest er at undersøge afstandsvariablen. Kontinuitets antagelsen tilsiger, at ansøgere ikke må kunne manipulere eller have perfekt kontrol over afstandsvariablen. Dette kan undersøges ved, at se på fordelingen af afstandsvariablen. (Lee & Lemieux, 2010) Hvis ansøgerne ikke har muligheden for præcis at manipulere med deres afstandsvariabel, så burde der ikke være en ophobning af ansøgere lige ved og over cut-off. Denne antagelse er ikke nødvendig for at RD designet skal holde, men resultaterne fremstår mere troværdige, hvis dette gør sig gældende. Denne form for test kaldes ofte en densitetstest. (Cattaneo, 2017) Figur 8.1 viser histogrammet med afstandsvariablen, hvor det kan ses, at fordelingen umiddelbart ikke er normalfordelt omkring 0, men i stedet har et midtpunkt omkring 1. Denne visuelle inspektion af afstandsvariablen tyder derfor på, at ansøgerne har muligheden for at manipulere med afstandsvariablen.

I figur 8.1 ser det ud til, at der forekommer en del sortering omkring cut-off. Således at individerne formår at sortere sig selv ind på højresiden af cut-off. Den visuelle undersøgelse tyder på, at der forekommer en manipulation, hvorfor jeg nu empirisk tester om dette er tilfældet. Jeg opstiller nulhypotesen, om at fordelingen af afstandsvariablen er kontinuiteret omkring cut-off, hvilket vil medføre, at der derfor ikke findes nogen manipulation nær cut-off. Testværdien vil i dette tilfælde være $-1,394$ og den tilhørende p-værdi er $0,1633$. Dette betyder, at jeg ikke kan forkaste nulhypotesen om, at der ikke er nogen forskel mellem fordelingen af behandlede og kontrol observation nær cut-off.

Hvis ansøgerne er i stand til at forudsige adgangskvotienten på den søgte ud-

Figur 8.1: Fordeling af afstandvariabel, 2016-2018



Note: Den massive linje viser normalfordelingen, mens den røde stiblet linje viser hvor afstanden til cut-off er lig 0.
Kilde: Egne beregninger baseret på registerdata fra Danmarks Statistik

dannelse i ansøgningsåret, så er ansøgerne som er under og over adgangskvotienten meget tilbøjelige til at være forskellige. Derfor kan det være relevante at undersøge de to elementer af afstandsvariablen, nemlig det adgangsgivende karaktergennemsnit fra en ungdomsuddannelse og grænsekvotienten, som angivet i ligning 6.2. Det adgangsgivende karaktergennemsnit fra en ungdomsuddannelse består af et vægtet gennemsnit af karakterne af eksamenskarakter for hvert fag ansøgeren har haft på sin ungdomsuddannelse. Disse karakter for hvert fag er sammenlignelige på tværs af ungdomsuddannelser idet, at disse er monitoreret af Undervisningsministeriet, som fastsætter de faglige krav for hvert enkelt fag. Derudover er alle skriftlige eksaminer ensartet på tværs af landet, og for større opgaver og mundtlige eksaminer vil blive bedømt af underviser og en ekstern censor. (Joensen & Nielsen, 2009) Derudover vil karaktergennemsnittet fra en ungdomsuddannelse ligge fast efter færdiggørelsen af en ungdomsuddannelse, og det vil ikke være muligt for den enkelte at ændre på dette efterfølgende. Dette karaktergennemsnit vil derfor bestemmes før at grænsekvotienterne til de videregående uddannelser bestemmes. Det vil derved være usandsynligt, at det er muligt for ansøgerne at manipulere med sit karaktergennemsnit ved ansøgningstidspunktet.

Dog er det fortsat muligt, at ansøgerne kan have perfekt kontrol, hvis det gør sig gældende, at ansøgerne perfekt kan forudsige adgangskvotienten for den søgte uddannelse i ansøgningsåret. En perfekt forudsigelse vil netop betyde, at ansøgerne har muligheden for at fravælge uddannelse, hvor deres karaktergennemsnit er lavere end grænsekquotienten. Dette vil medføre, at ansøgerne over og under cut-off sandsynligvis vil være forskellige, hvilket bryder med kontinuitetsantagelsen. Denne hændelse kan forekomme hvis en uddannelse gentagne år har den samme grænsekquotient, hvorfor variationen i denne vil være meget lille. Det må antages, at ansøgerne bruger tidligere års adgangskvotienter til at forudsige adgangskvotienten i ansøgningsåret, hvorfor en lav variation i grænsekquotienten vil give ansøgerne muligheden for relativt præcist at forudsige adgangskvotienten.

Derfor er det relevant at vide, hvor meget at grænsekquotienterne for hvert uddannelsesudbud har ændret år til år. I det anvendte data har der i den undersøgte periode været udbudt 189 forskellige videregående uddannelsesudbud, hvor der har en grænsekquotient i den undersøgte periode fra 2016-2018. Det store udbud kan derfor gøre det svært at vurdere ændringen i grænsekquotienterne fra år til år. Derfor undersøges den gennemsnitlige absolutte ændring og standardafvigelsen for den absolutte ændring i grænsekquotienterne fra år til år for hvert uddannelsesudbud. (Humlum et al., 2012) Den gennemsnitlige numeriske ændring er 0,171 med en standard afvigelse på 0,236, hvorfor det må påstås, at der har været en rimelig variation i grænsekquotienterne fra år til år i den undersøgte periode. Dog er antallet af ansøgere langt fra ens for hvert uddannelsesudbud, hvorfor det vil være retvisende at finde det vægtede gennemsnit af den numeriske ændring i grænsekquotienten, hvor vægtene vil afspejle antallet af ansøgerne i det anvendte datasæt. Her opnås et vægtet gennemsnit på 0,121 med en standard afvigelse på 0,176. Dette gennemsnit er altså noget lavere end det ikke vægtede gennemsnit, hvilket kan forklares med, at størstedelen af ansøgerne findes på uddannelsesudbud med relativ lav variation i grænsekquotienterne.

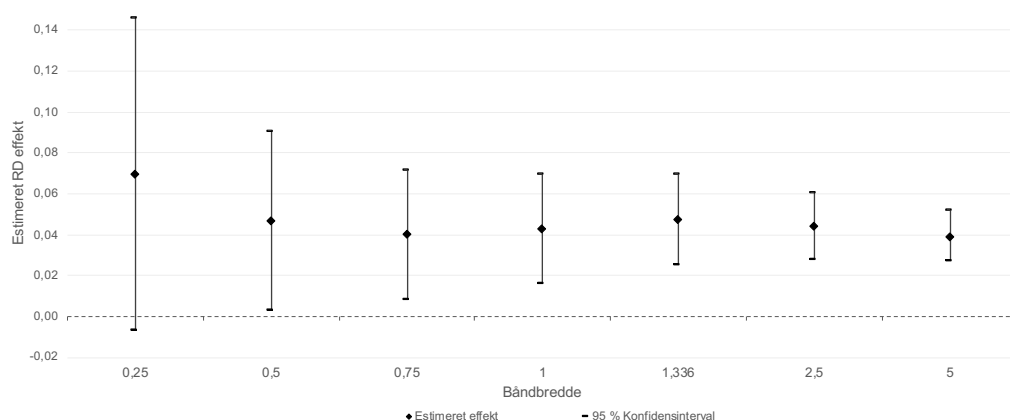
Det skal desuden bemærkes, at i udregningen af den numeriske ændring fra år til år ikke medtager alle uddannelser i årene. Dette skyldes, at for nogle uddannelsesudbud har en grænsekquotient i det ene år, mens i det efterfølgende år ikke har nogle. Der vil således ikke være noget mål for den numeriske ændring. Der har været en rimelig variation i den gennemsnitlige vægtede numeriske ændring i grænsekvo-

tienterne, hvorfor det må formodes, at potentielle ansøgere ikke vil være i stand til perfekt at forudsige adgangskravet i det år, de ansøger. Derudover skal det bemærkes, at ændringerne på tværs af uddannelsesudbuddene ikke har været samme retning. For nogle udbud har der forekommet en positiv ændring i grænsekvisitienten, mens for andre har der været en negativ ændring. Alt i alt tyder dette på, at der har været en rimelig variation i grænsekvisitienterne til, at ansøgerne ikke perfekt kan forudsige grænsekvisitienten i ansøgningsåret.

8.2 Båndbreddens betydning

En anden væsentlig og vigtig validitetstest er at undersøge, hvor sensitiv resultaterne er overfor valget af båndbredde. Som tidligere nævnt i afsnit 6 vil der opstå et trade-off mellem præcise estimater og bias, når båndbredde vælges. Den MSE optimale båndbredde forsøger at minimere dette trade-off problem og vælger ud fra en data drevet metode den rette båndbredde. En undersøgelse af sensitibiliteten vil derfor her blive illustreret efter fremgangsmåden af Lee & Lemieux, 2010, hvor der grafisk præsenteres et plot af estimaterne med et 95% konfidensinterval for hver båndbredde. Figur 8.2 viser modellen i ligning 6.3, hvor der er inkluderet fixed effects på uddannelsesudbud.

Figur 8.2: Sensitiviteten af estimater i forhold til valget af båndbredden



Note: Den anvendte model har fixed effects på uddannelsesudbud. Båndbredden på 1,336 er den MSE optimale båndbredde

Kilde: Egne beregninger baseret på registerdata fra Danmarks Statistik

Figuren i 8.2 viser, at resultaterne er meget lidt sensitiv overfor valget af båndbredden. Effekten af optagelseskvisitient på frafaldssandsynligheden varierer ganske

lidt på tværs af valget af båndbredden. Dog kan det ses, at estimatet med en båndbredde på 0,25 er relativt større end de resterende estimater, hvilket heller ikke er særlig overraskende, da der anvendes relativt få observation for at opnå dette resultat, således at få observation tillægges stor betydning. Det bør dog bemærkes, at dette estimat er insignifikant. Bemærkelsesværdigt er det dog, at de øvrige estimerede RD estimater er i nogenlunde samme størrelsesorden, som resultaterne opnået med den MSE optimale båndbredde, hvilket er med til at verificerer designet og resultaterne. Resultater rimelig stabile og robuste overfor en ændring i valget af båndbredde, hvilket bestyrker robustheden af de opnåede resultater i det forrige afsnit. (Lee Lemieux, 2010)

8.3 Forudbestemte kovariater

En anden, men meget relevant, metode at teste robustheden af resultaterne er ved at undersøge kovariaterne. En af de vigtigste falsifikation test af RD designet er, at undersøge om individerne tæt på cuf-off ligner hinanden med hensyn til observerbare karakteristika.

Idéen bag denne test er, at hvis individerne ikke har præcis mulighed for at manipulere med deres afstandsvariabel, så burde der ikke være nogle systematisk forskelle mellem individer, som opnår behandling og dem der ikke gør. Forudbestemte kovariater bør i følge antagelserne bag RD designet ikke blive påvirket af behandlingen, hvorfor at hvis kovariater korrelerer stærkt med frafaldssandsynligheden og er diskontinuerlige ved cuf-off vil RD designet være tvivlsomt, og resultaterne vil ikke holde. For at teste om dette er tilfældet foretages der en RD regression for hver af disse kovariater. I tabel 8.1 kan disse regressioner ses.

Tabel 8.1: Test af kovariat

Variable	MSE-optimal båndbredde	RD estimat	P-værdi
Køn	1,479	0,008	0,621
Alder	1,384	-0,274	0,116
Herkomst	1,219	0,049	0,320
Mor har en lang videregående uddannelse	1,181	-0,022	0,151
Far har en lang videregående uddannelse	1,181	-0,021	0,271

Tabel 8.1 viser punktestimater for RD regressionerne for hvert forudbestemte kovariater. Det kan ses i tabellen at disse estimater er små, og de fleste er tæt på

nul. Dog med undtagelse af alder variabelen. Men alle variable er ikke signifikante, da alle p-værdier er i intervallet fra 0,116-0,621. Derfor findes der ikke empiriske bevis for, at disse forudbestemte kovariater er diskontinuerte ved cut-off. Det bør dog bemærkes, at antallet af observationer som er brugt til hver regression for kovariater er forskellige. Dette skyldes, at den optimale MSE-båndbredde er forskellig for hver af disse kovariater. (Lee Lemieux, 2010)

9 Diskussion

I dette afsnit diskuteres de opnåede resultater, samt metoden herfor. Mine resultater tyder på, at studerende, der er optaget i kvote 2, har en lavere frafaldssandsynlighed i løbet af det første studieår. Den eksterne validitet blive belyst i dette afsnit, samt yderligere metodiske begrænsninger af de opnåede resultater. Derudover vil de politiske interventioner som følge af resultaterne vil blive .

9.1 Begrænsninger

I dette afsnit diskuteres nogle af de metodiske begrænsninger ved dette speciale. Effekterne er lokale gennemsnitlige behandlingseffekter, hvorfor det er relevant at diskutere deres eksterne validitet. Ligeledes vil metodiske begrænsninger for dette speciale også blive diskuteret nedenfor.

9.1.1 Validitet

De opnåede resultater fra afsnit 7 er lokale gennemsnitlige behandlingseffekter, som tyder på en høj intern validitet, og således troværdige inden for dette speciales egne metodiske præmisser. En og måske den største begrænsning af RD designet er manglen på ekstern validitet. I mange tilfælde fører designet til estimater med en høj grad af intern gyldighed, mens den eksterne halter efter. Det er dog alligevel værd, at diskutere hvorvidt disse resultater også er eksterne valide og mulige at generalisere til andre situationer, tidspunkter, eller til andre befolkningsgrupper end for det anvendte sample.

I RD designet anvendes kun observationer nær cut-off, som beskrevet i afsnit 6, hvorfor de estimerede resultater karakteriseres som lokale behandlingseffekt i diskontinuitetspunktet. Hvorfor RD designet kun i stand til at estimere en lokal kausal effekt, der gælder for observationer omkring cut-off. (Imbens og Lemieux, 2008) Dette betyder, at den høje interne validitet, som RD designet er forbundet med, og de få antagelser har den omkostning, at estimatet af den kausale effekt er begrænset til den subpopulation af data, som ligger omkring cut-off værdien.

Derfor er udfordringen her at beskrive, i hvilket omfang estimatet af den kausale effekt kan forventes at se ud for observationer længere væk fra cut-off værdien. Hvad kan der forventes, at der vil ske med den kausale effekt i takt med, at vi bevæger os væk fra cut-off værdien? Man bør altid overveje om særlige forhold gør sig gældende

for observationer omkring cut-off værdien. Dette er dog ikke væsentligt forskelligt fra de fleste eksperimentelle designs, hvor estimater altid vil være lokale i forhold til en særlig population, kontekst eller tid. (Angrist og Pischke, 2009).

For dette speciale må det formodes, at de opnåede effekter af optagelseskvoten på de studerendes frafaldssandsynlighed vil have samme fortegn for observationer længere væk fra cut-off. Hvilket kan forklares med, at det er optagelsen i kvote 2 giver de studerende et ekstra 'boost', som styrker deres vedholdenhed. Det kan endda formodes, at studerende, som er optaget i kvote 2 og er langt fra cut-off oplever et endnu stærkere 'boost' og har en lavere frafaldssandsynlighed. En forklaring på dette skyldes, at kvote 2 studerende har typisk ringere muligheder for at søge andre uddannelser, da de har et lavere karaktergennemsnit, hvilket betyder, at deres alternative uddannelsesmuligheder er begrænset, da disse har et lavt karaktergennemsnit fra en ungdomsuddannelse.

9.1.2 Metodiske begrænsninger

I robusthedcheckene af de forudbestemt kovariater i tabel 8.1 kan det aflæses, at RD estimatet for alder har en p-værdi på 0,116, hvorfor denne er meget tæt på, at være signifikant på et 10% niveau. Det kan således ikke afvises, at der er en diskontinuitet i denne baggrundsvARIABLE, hvilket betyder at der vil være en forskel på ansøger på hver side af tærskelværdien. Resultaterne er relative robuste i forhold til de udførte robusthedscheck.

Dog viser analyserne af baggrundvariablene, at der kan være et potentielt selektionsproblem som følge af en begrænset variation i grænsekvotienterne. Den begrænsede variation i grænsekvotienter kan bl.a. skyldes, at den undersøgte periode er relativ snæver og består af 3 optagelsesår. Det er muligt og sandsynligt, at variationen i grænsekvotienterne havde været større, hvis en længere tidsperiode havde været anvendt.

Derudover er det praksis ikke muligt kun at vælge data helt tæt på cut-off. Dette skyldes, at des tættere det undersøgte område bliver, des færre datapunkter vil der være. Det kan betyde, at der ikke er nogle observationer overhovedet. Derfor træffes valget om, at også benytte observationer fra cut-off. (Lee & Lemieux, 2010) Derfor er der i dette speciale anvendte en MSE optimal båndbredde, samt de opnåede

resultater er udelukkende lokale effekter, som nævnt ovenfor.

Disse metodiske udfordringer og begrænsninger betyder, at validiteten af resultaterne kan være forbundet med usikkerhed, og derved kan det være muligt, at disse ikke er generaliserbare.

9.1.3 Justification bias

Det er muligt at diskutere, om inklusionen af det selvrapporterede niveau af afklarethed er retvisende og relevant at anvende idet, at måle tidspunktet for denne variabel finder sted efter afgørelsen af, om den enkelte studerende er optaget i kvote 1 eller kvote 2. Det kan derfor argumenteres, at denne variabel kan være påvirket af justification bias, og estimerne hermed også vil være biased.

Justification bias kan opstå ved at sammenligne variationen mellem de to mål inden for enkelt personer over tid med variationen frafaldssandsynligheden over tid. Derfor anvendes denne variabel også kun som et robusthedscheck for at se, om størrelsesordenen på estimatet fortsat vil være den samme. (Angrist & Pischke, 2009)

9.2 Politiske interventioner

Frafald i løbet af første studieår har både negative konsekvenser for uddannelsesinstitutionernes økonomi og for den enkelte studerende. Frarafald er uden tvivl et helt fundamentalt uddannelseseffektmål, som bør tages seriøst for alle interessenters skyld, hvorfor det er værd, at overveje mulige politiske interventioner som følge af resultaterne.

9.2.1 Bør kvote 2 udvides?

Resultaterne i afsnit 7 ligger naturligt op til spørgsmålet om kvote 2 udvides? Resultaterne indikerer netop, at hvis man ønsker, at mindske frafaldet på de danske videregående uddannelser kan kvote 2 afhjælpe dette. De heterogene effekter viser ligeledes, at det er særlige grupper af studerende, som oplever et ekstra 'boost' i deres vedholdenhed af at være optaget i kvote 2. Det er endnu ikke klart, hvad der i kvote 2 udgør en frafaldsbeskyttende effekt. En formodning kan være, at disse studerende er oplever, at de via optagelse i kvote 2 'overlever' optagelsesprøverne eller muligvis er mere motiveret. I dette speciale er der inkluderet mål for selvrapporteret afklarethed, som tydeligt indikerer, at studerende optaget i kvote 2 angiver et højere niveau af

afklarethed sammenlignet med studerende, som er optaget i kvote 1. Men det er muligt, at selve optagelsesproceduren også er med til at påvirke kvote 2 optagnes efterfølgende studiemotivation.

Danmarks Evalueringsinstitut foretog i 2017 en undersøgelse af, hvilken effekt at optagelsessamtalerne på læreruddannelsen har. Denne uddannelse har nemlig siden 2013 haft obligatoriske optagelsessamtaler, hvis ansøgeren har et karaktergennemsnit under 7,0. Undersøgelsen viser, at optagelsessamtalerne på uddannelsen har bidraget til at mindske frafaldet blandt studerende, som er optaget via kvote 2. Studerende, som er optaget i kvote 1, estimeres til, at have en frafaldssandsynlighed på 21,4%, men studerende, som er optaget i kvote 2, estimeres til, at have en frafaldssandsynlighed på 16,8%. Dette svarer til en effekt på -4,6%. (Danmarks Evalueringsinstitut, 2017b) Størrelsesordenen på denne effekt er meget tilsvarende den effekt, som er fundet i dette speciale, selvom der i dette speciale udelukkende anvendes universitetsbachelorstuderende. Resultaterne indikerer, at det er muligt, at nedbringe det tidlige frafald for studerende, hvis flere optages i kvote 2.

Resultaterne fra dette speciale indikerer, at en udvidelse af kvote 2 optaget kan være med til at nedbringe det samlede tidlige studiefrafald. Det er dog muligt, at hvis denne kvote udvides markant vil det 'boost' som optagne i kvote 2 oplever mindske, da oplevelsen af at komme gennem nåleøjet mindskes. Det formodes dog, at den ekstra tid som studerende, som er optaget i kvote 2, medvirker at disse bliver ekstra afklaret og mere motiveret, som indikerer, at det kan være en mulighed at udvides kvote 2 optaget for at minimere frafaldet.

Det er dog værd at nævne, at man ikke altid kan gå ud fra at det uhen-sigtsmæssigt at studerende afbryder deres uddannelse før tiden, som både vedrører dannelsesmæssige over for økonomiske hensyn. Desuden kan en noget af studiefrafaldet i højere grad karakteriseres som et omvalg fremfor et reelt frafald. Det er imidlertid der findes dog ikke et kortlægning af, hvor stort omvalgsandelen er for universitetsbachelor i samme periode, som dette speciale. Fra et økonomisk perspektiv er det uhensigtsmæssigt at de studerende og universiteterne bruger tid og ressourcer på uddannelser de studerende ikke fuldfører, hvorfor det er et politisk ønske at mindske frafaldet.

9.2.2 Er det danske optagesystem indrettet på en retfærdig og optimal måde?

Den sociale retfærdighed og mobilitet er også et element i kvotesystemet, da kvote 2 anses for at give muligheden for de ansøgere, og således kan omtales som en "second chance". Det bør dog alligevel overvejes, om det gennem dette kvotesystemet er muligt, at rykke den sociale mobilitet væsentlig, da der allerede er sket en selektion af ansøgerne langt forinden ansøgningen til de videregående uddannelser. (Uddannelses- og forskningsministeriet, 2020) Som beskrevet i afsnit 4 så har tidligere undersøgelser vist, at studerende af forældre med en lang videregående uddannelse får en forlomme både når det gælder om, at opnå høje karakterer fra en ungdomsuddannelse og blive optaget i kvote 2. Det er dermed begrænset, hvor meget at det er muligt, at øge den sociale diversitet blandt kvote 2-optagne i forhold til kvote 1. (O'Neill et al., 2011)

Væsentligt er det dog at have for øje, at det danske optagelsessystem er et komplekst system, hvilket kan gøre det svært for den enkelte unge, når denne skal træffe en beslutning om at søge en uddannelse. Der findes et meget stort udbud af uddannelser i Danmark, hvilket fodrer at forudsætninger til, hvad det kræver for optagelse på uddannelsen er mest mulig gennemskuelige. Dette betyder også, at man bør være varsom med at ændre for meget i det nuværende system. En af de bedste prædikater for studerendes gennemførsel og akademiske præstationer er karaktergennemsnittet fra en ungdomsuddannelse. Her vil høje karakter fra en ungdomsuddannelse fører til succes på studiet i form af lavere frafald, lav dimittendledighed og faglig succes. Dette bør kraftig overvejes, hvis man ønsker et optagelsessystem, som har differentierede krav. Disse kan nemlig være med til at mudre systemet, og gøre det mere besværligt for den unge at vælge uddannelse. (Danmarks Evalueringsinstitut, 2015)

9.3 Mulige fremtidige undersøgelser

Der er publiceret ganske få studier, som har beskæftiget sig med hvilken effekt, at kvote 2 har på de studerendes frafaldssandsynlighed, hvorfor dette speciales eksperimentelle indstilling til tilgangen for at besvare undersøgelsesspørgsmålet bidrager med viden til området. Det er dog fortsat uklart, hvad det præcist er, som kan forklare hvorfor kvote 2 har en frafaldsbeskyttende effekt. Derfor vil en mere dybdegående undersøgelse af, hvilke elementer i kvote 2 mindske sandsynligheden for frafald selvfølgelig stadig påkrævet og meget interessant. En sådan undersøgelse vil kunne være med til at understøt

Ligeledes vil en tilsvarende undersøgelse af betydningen af optagelseskvoten på frafaldssandsynligheden for en længere periode være interessant idet, at denne vil være med til at forstærke de fundne effekter. Derudover vil det være interessant at undersøge, om de fundne kausale effekter for universitetsbacheloror også gør sig gældende for de studerende på professionsskolerne og erhvervsakdemierne.

10 Konklusion

Dette speciale forskningsspørgsmål var at undersøge, om der er effekter af optagelseskvoten på de studerendes frafaldssandsynlighed. Med anvendelse af et regression diskontinuitet design, som gør det muligt, at opnå kausale effekter, så længe af antagelserne er opfyldt, kan dette forskningsspørgsmål besvares. Resultaterne fra dette speciale viser netop, at der er en forskel i frafaldssandsynligheden på 4.8 pct. point mellem studerende optaget i kvote 1 sammenlignet med kvote 2. I dette estimat er der desuden kontrolleret for uddannelsesspecifikke fixed effects, og derved minimeres variationen fra udeladte variable. Resultaterne giver ikke en forklaring på, hvad det præcist er, som kan forklare hvorfor kvote 2 har en frafaldsbeskyttende effekt. En mulig årsag kan være, at optagelsesredskaberne i kvote 2 er bedre til at matche ansøger og uddannelser, samt selve optagelsesprocessen giver den studerende et form for 'boost' som øger motivationen.

En undersøgelse af de heterogene effekter viser derudover, at det særligt er optagne mænd i kvote 2 og optagne i kvote 2 med mindst en forælder med en lang videregående uddannelse, som driver denne effekt. Disse typer af studerende oplever et særligt 'boost' ved optagelse i kvote 2, og dermed falder disse mindre fra sammenlignet med samme grupper, som er optaget i kvote 1.

Resultater fra dette speciale indikerer, at en mulig intervention kan være, at udvide optaget i kvote 2, da dette overordnet kan være med til at mindske det samlede studiefrafald, som finder sted i løbet af det første studieår. Det er dog værd at bemærke, at de opnåede effekter er lokale, og derfor bør man være varsom med den eksterne validitet af resultaterne, som også angivet i diskussionen.

11 Litteraturliste

Angrist, J. D. & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Becker, G. (1964), *Human capital: A theoretical and empirical analysis, with special reference to education*. Chicago, University of Chicago Press, UCP descry.

Bingley, P., Christensen, K. & Walker, I. (2009). *The Returns to Observed and Unobserved Skills over Time: Evidence from a Panel of the Population of Danish Twins*.

Card, D. (1999) *Part A of Handbook of Labor Economics* , (pp. 1801 – 1863). Elsevier.

Cattaneo M. D. & Escanciano J. C. (2017) *Regression Discontinuity Designs: Theory and Applications*, Advances in Econometrics, volume 38, Emerald Group Publishing

Danmarks Evalueringsinstitut (2015), *Den sociale profil i optagesystemet*

Danmarks Evalueringsinstitut (2017a). *Studiestartens betydning for frafald på videregående uddannelser* .

Danmarks Evalueringsinstitut (2017b) *Effekten af optagelsessamtaler på læreruddannelsen*.

Danmarks Evalueringsinstitut (2019), *Kortlægning af kvote 2-systemet på de videregående uddannelser*

Dansk Erhverv (2019) *Pigerne er bedst til matematik i gymnasiet*

DEA (2014), *Stort set ingen forskel på frafald i kvote 1 og kvote 2*. Notat 34

Humlum M. K., Kleinjans K. J. & Nielsen H. S (2012) *An economic analysis of identity and career choice*, Economic inquiry, 2012

Imbens, G. & Kalyanaraman, K. (2012). *Optimal bandwidth choice for the regression discontinuity estimator*. Review of Economic Studies, 79(3), 933 – 959.

Imbens, G. W. & Lemieux, T. (2008). *Regression discontinuity designs: A guide to practice*. Journal of Econometrics, 142(2), 615 – 635. The regression discontinuity

design: Theory and applications.

Imbens, G. & Kalyanaraman, K. (2012). *Optimal bandwidth choice for the regression discontinuity estimator*. *Review of Economic Studies*, 79(3), 933 – 959.

Ishitani, T.T. (2003) *A Longitudinal Approach to Assessing Attrition Behavior Among First-Generation Students: Time-Varying Effects of Pre-College Characteristics*. *Research in Higher Education* 44, 433–449

Joensen, J. S. & Nielsen, H. S. (2009). *Is there a causal effect of high school math on labor market outcomes?* *Journal of Human Resources*, 44(1), 171–198.

Lee, D. S. & Lemieux, T. (2010). *Regression discontinuity designs in economics*. *Journal of Economic Literature*, 48(2), 281–355.

Li, F. Mattei, A. & Mealli, F (2015). *Evaluating the causal effect of university grants on student dropout: Evidence from a regression discontinuity design using principal stratification*. *Ann. Appl. Stat.* 9, no. 4, 1906–1931

OECD (2020) *Education at a Glance 2020 - OECD INDICATORS*

O'Neill, L. D., Vonsild, M., Wallstedt, B. (2011). *Could non-grade based selection improve medical student socio-demography diversity?* I: Medical Education (in review).

Skovron C. & Tituni R. (2015) *A practical guide to regression discontinuity designs in political science*, *American Journal of Political Science*, 2015

Spence, M. (1973). *Job market signaling*. *Quarterly Journal of Economics*, 87(3), 355 – 374.

Studenterrådet ved Aarhus Universitet. (2000) *Frafald og studiemiljø*

Tinto, Vincent. (1993). *Leaving college: rethinking the causes and cures of student attrition*, 2nd ed. Chicago; London: University of Chicago Press.

Tinto, Vincent. (2007). *Research and Practice of Student Retention: What Next?*, *Journal of College Student Retention*; London 8(1):1–19.

Thistlethwaite, D. L., Campbell, D. T. (1960). *Regression-discontinuity analysis: An alternative to the ex post facto experiment*. *Journal of Educational Psychology*, 51(6), 309–317

Thomsen, Jens Peter. (2016) *Test-based admission to selective universities: A lever for first-generation students or a safety net for the professional classes?*, VIVE

Troelsen, Rie. (2011). *Frafald på de videregående uddannelser – hvad ved vi om årsagerne?* Dansk Universitetspædagogisk Tidsskrift 6(10):37–44.

Uddannelse- og forskningsministeriet. (2017) *Analysenotat – frafald på de videregående uddannelser*

Uddannelse- og forskningsministeriet. (2018) *Frafald og studieskift*

Udvalg for Kvalitet og Relevans i de Videregående Uddannelser. (2013) *Nye veje og høje mål*

Wooldridge, J. M. (2009) *Econometric analysis of cross section and panel data*. MIT press.

Appendiks A: Data

Spørgeskema fra Danmarks Evalueringsinstitut fra første bølge

Nedenfor vises indholdet af det fulde spørgeskema, som er udsendt til studerende, som har søgt om optagelse på en lang videregående uddannelse. De studerende har modtaget spørgeskemaet på e-boks i august måned.

```
<On Error GoTo Default_ErrorHandler>
[If IOM.Info.IsDebug = False Then]
[BEGIN LOOP – ClientInfo (repeat 5 times)]
Client Info
1. (UserAgent) User Agent
2. (Browser) Browser
3. (OS) OS
4. (DeviceType) Device Type
5. (DeviceModel) Device Model

[END LOOP – ClientInfo]
[If LCase(IOM.Info.User4) = "refuse" Then Goto TerminateScriptBM]
[If (IOM.Info.IsDebug = "True") Then]
    <GoTo Here>
[End If]
```

1.1.1.1 <Here:>

[Intro – INFO – single – Must answer]

Tak fordi du vil deltage i vores undersøgelse. I det første spørgeskema sætter vi fokus på tiden inden studiestart: Hvilke forventninger har du, hvad motiverer dig, og hvilke mål har du med uddannelsen? Vi ser frem til at få dine svar. Klik på "Næste" for at begynde din besvarelse. Med venlig hilsen Jakob Rathlev Områdechef, Danmarks Evalueringsinstitut

[BEGIN PAGE – q1_page]

[q1 – CATEGORICAL – single – Must answer]

Forventer du at starte på den uddannelse, du er blevet optaget på?

- ☐ (1) Ja
- ☐ (2) Nej

[END PAGE – q1_page]

[If q1 = (2) Then]

[q3 – GRID – single – Must answer - Randomized]

I hvilken grad har følgende været medvirkende til, at du ikke starter på den uddannelse, du er blevet optaget på? {#ex_script}

	I høj grad	I nogen grad	I mindre grad	Slet ikke	Ved ikke
Jeg forventer at starte på en anden uddannelse, end den jeg er kommet ind på	☐	☐	☐	☐	☐
Jeg har fundet ud af, at uddannelsen er anderledes, end jeg troede	☐	☐	☐	☐	☐
Det var ikke en velovervejet beslutning, at jeg søgte ind på uddannelsen	☐	☐	☐	☐	☐

Uddannelsesindsatsen i Danmark 2019

Jeg ønsker alligevel ikke at starte på en uddannelse nu	☐	☐	☐	☐	☐
Alvorlige livsbegivenheder (fx dødsfald, sygdom, brud med kæreste/kone)	☐	☐	☐	☐	☐
Jeg er blevet tilbudt et job/fortsætter i et job, som jeg foretrækker frem for uddannelsen	☐	☐	☐	☐	☐
Praktiske forhold forbundet med at begynde på uddannelsen (fx bolig, flytning eller andet)	☐	☐	☐	☐	☐
Økonomiske problemer	☐	☐	☐	☐	☐

[q4 – CATEGORICAL – single – Must answer]

Kan du uddybe årsagen til, at du ikke starter på den uddannelse, du er blevet optaget på?

- ☐ (1) Ja, notér:
- ☐ (2) Nej

<Goto TerminateScriptBM>

[End If]

[q5 – CATEGORICAL – single – Must answer]

Har du på nuværende tidspunkt deltaget i et introduktionsforløb/studiestartsforløb/rustur/rusforløb eller lignende i forbindelse med den uddannelse, du er blevet optaget på?

- ☐ (1) Ja
- ☐ (2) Nej

[q6 – CATEGORICAL – single – Must answer]

Hvornår traf du beslutningen om, at du ville søge ind på den uddannelse, du er blevet optaget på?

- ☐ (1) Dagene op til ansøgningsfristen
- ☐ (2) Månedene op til ansøgningsfristen
- ☐ (3) Mellem én måned og et halvt år før ansøgningsfristen
- ☐ (4) Mellem et halvt år og to år før ansøgningsfristen
- ☐ (5) Mere end to år før ansøgningsfristen
- ☐ (6) Ved ikke

[q7 – GRID – single – Must answer]

I hvilken grad passer følgende udsagn på dig: {#ex_script}

	I meget høj grad	I høj grad	I nogen grad	I mindre grad	Slet ikke	Ved ikke
Da jeg søgte, følte jeg mig helt afklaret med, at jeg ville studere på netop den	☐	☐	☐	☐	☐	☐

uddannelse, jeg nu er blevet optaget på						
Da jeg søgte, havde jeg tilstrækkelig information om den uddannelse, jeg nu er blevet optaget på	☐	☐	☐	☐	☐	☐

[q8 – GRID – single – Must answer]

{#ex_script}

	Ja	Nej	Ved ikke
Er du blevet optaget på den <i>uddannelsesinstitution</i> (fx Københavns Universitet eller Erhvervsakademi Aarhus), du helst vil ind på?	☐	☐	☐
Er du blevet optaget på den <i>uddannelse</i> (fx psykologi eller læreruddannelsen), du helst vil ind på?	☐	☐	☐

[q9 – GRID – single – Must answer - Randomized]

I hvilken grad var følgende forhold afgørende for din beslutning om at søge ind på den uddannelse, du er blevet optaget på?

{#ex_script9}

	Slet ikke	I mindre grad	I nogen grad	I høj grad	I meget høj grad
Min faglige interesse for uddannelsen	☐	☐	☐	☐	☐
Det sociale miljø på uddannelsen	☐	☐	☐	☐	☐
Den geografiske placering af uddannelsen	☐	☐	☐	☐	☐
Jeg ønskede at blive en mere dannet/vidende person	☐	☐	☐	☐	☐
Jeg ville gøre min familie glad og stolt	☐	☐	☐	☐	☐
Jeg blev anbefalet uddannelsen af en bekendt	☐	☐	☐	☐	☐
Prestige forbundet med uddannelsen	☐	☐	☐	☐	☐

Gode beskæftigelsesmuligheder	☐	☐	☐	☐	☐
Høj indkomst	☐	☐	☐	☐	☐
Jeg ønskede at øge mine chancer for senere at komme ind på en anden uddannelse	☐	☐	☐	☐	☐

[q10A – CATEGORICAL – single – Must answer]

Hvor motiveret er du for at starte på uddannelsen?

- ☐ (1) 0 – Slet ikke motiveret
- ☐ (2) 1
- ☐ (3) 2
- ☐ (4) 3
- ☐ (5) 4
- ☐ (6) 5
- ☐ (7) 6
- ☐ (8) 7
- ☐ (9) 8
- ☐ (10) 9
- ☐ (11) 10 – Ekstremt motiveret
- ☐ (12) Ved ikke

[q10B – GRID – single – Must answer - Randomized]

Du bedes nu bedømme følgende udsagn på baggrund af dine forventninger til, hvordan du vil klare dig på den uddannelse, du er blevet optaget på. Du skal bruge en 7-punkts-skala, hvor 1 betyder, at udsagnet <i>slet ikke passer</i> på dig, og hvor 7 betyder, at udsagnet <i>passer præcist</i> på dig: {#ex_script7}

	1 Passer slet ikke	2	3	4	5	6	7 Passer præcist
Jeg tror, at jeg vil få rigtig gode karakterer på uddannelsen	☐	☐	☐	☐	☐	☐	☐
Jeg er overbevist om, at jeg kan forstå selv det sværeste stof på uddannelsen	☐	☐	☐	☐	☐	☐	☐
Jeg er sikker på, at jeg kan forstå de grundlæggende begreber på uddannelsen	☐	☐	☐	☐	☐	☐	☐
Jeg er overbevist om, at jeg kommer til at klare mig rigtig godt i opgaver og til prøver på uddannelsen	☐	☐	☐	☐	☐	☐	☐

Jeg forventer, at mine evner som studerende er fremragende sammenlignet med andre, der er kommet ind på uddannelsen	☐	☐	☐	☐	☐	☐	☐
---	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

[q11 – GRID – single – Must answer - Randomized]

Nedenfor præsenteres nogle udsagn. Der er ikke noget rigtigt eller forkert svar. Svar på baggrund af dine egne følelser, i stedet for hvordan du tænker, de fleste andre ville svare. I hvilken grad passer følgende udsagn på dig: {#ex_script}

	I meget høj grad	I høj grad	I nogen grad	I mindre grad	Slet ikke	Ved ikke
I perioder præget af uvished, forventer jeg som regel det bedste	☐	☐	☐	☐	☐	☐
Hvis noget kan gå galt for mig, så går det galt	☐	☐	☐	☐	☐	☐
Jeg er altid optimistisk omkring min fremtid	☐	☐	☐	☐	☐	☐
Jeg regner sjældent med, at der sker noget godt for mig	☐	☐	☐	☐	☐	☐

[q12 – GRID – single – Must answer - Randomized]

Du bedes bedømme, i hvilket omfang følgende udsagn passer på dig. Du skal bruge en 7-punkts-skala, hvor 1 betyder, at udsagnet <i>slet ikke passer</i> på dig, og hvor 7 betyder, at udsagnet <i>passer præcist</i> på dig: {#ex_script7}

	1 Passer slet ikke	2	3	4	5	6	7 Passer præcist
Det mest tilfredsstillende ved at studere er for mig at opnå gode karakterer	☐	☐	☐	☐	☐	☐	☐
Hvis jeg kan, vil jeg gerne opnå bedre karakterer end de fleste andre	☐	☐	☐	☐	☐	☐	☐
Jeg vil gerne klare mig godt på min uddannelse, fordi det er vigtigt at demonstrere mine evner	☐	☐	☐	☐	☐	☐	☐

overfor min familie, mine venner eller andre							
Jeg foretrækker undervisningsmateriale, der er virkelig udfordrende for mig, så jeg kan lære noget nyt	☐	☐	☐	☐	☐	☐	☐
Jeg foretrækker undervisningsmateriale, der vækker min nysgerrighed, selvom det er svært at lære	☐	☐	☐	☐	☐	☐	☐
Det mest tilfredsstillende for mig er at forsøge at forstå indholdet i et fag så grundigt som muligt	☐	☐	☐	☐	☐	☐	☐
Når noget er svært, giver jeg enten op, eller jeg laver kun de nemme dele	☐	☐	☐	☐	☐	☐	☐
Selv når noget er kedeligt og uinteressant, bliver jeg ved med at arbejde med det, til jeg er færdig	☐	☐	☐	☐	☐	☐	☐

[BEGIN PAGE – q14_page]

[q14_intro – INFO – single – Must answer]

Bedøm venligst følgende udsagn på en 4-punkts-skala, hvor 1 betyder, at du <i>er dårlig til</i> det beskrevne og 4 betyder, at du <i>er god til</i> det beskrevne:

[q14_1 – CATEGORICAL – single – Must answer]

At finde på ting at lave sammen med nye mennesker, som jeg synes er interessante og attraktive

- ☐ (1) Jeg er <i>dårlig</i> til det her
- ☐ (2) Jeg er <i>kun nogenlunde</i> til det her
- ☐ (3) Jeg er <i>god</i> til det her
- ☐ (4) Jeg er <i>ekstremt god</i> til det her

[q14_2 – CATEGORICAL – single – Must answer]

At præsentere mig selv for nogen, som jeg gerne vil lære bedre at kende

- ☐ (1) Jeg er <i>dårlig</i> til det her
- ☐ (2) Jeg er <i>kun nogenlunde</i> til det her
- ☐ (3) Jeg er <i>god</i> til det her
- ☐ (4) Jeg er <i>ekstremt god</i> til det her

[END PAGE – q14_page]

[q15 – CATEGORICAL – single – Must answer]

Hvor mange timer forventer du ca. at bruge på at studere på den uddannelse, du er blevet optaget på (både undervisning og forberedelse) i løbet af en gennemsnitlig uge (syv dage)?

- ☐ (1) Notér antal timer:
- ☐ (2) Ved ikke

[q16 – CATEGORICAL – single – Must answer]

Hvad er det højeste uddannelsesniveau, du forventer at opnå i et livslangt perspektiv?

- ☐ (1) Erhvervsakademiuddannelse
- ☐ (2) Professionsbachelorgrad
- ☐ (3) Bachelorgrad
- ☐ (4) Kandidatgrad
- ☐ (5) Ph.d.-grad eller højere
- ☐ (6) Andet
- ☐ (7) Ved ikke

[BEGIN PAGE – q17_page]

[q17_intro – INFO – single – Must answer]

Til sidst vil vi gerne stille nogle få spørgsmål om dig selv og din nuværende livssituation:

[q17 – GRID – single – Must answer]

Vi er interesserede i, hvordan du har det med følgende udsagn. Udsagnene kan forekomme overlappende, men belyser forskellige aspekter. Læs hvert udsagn grundigt. Angiv hvordan du har det med hvert udsagn. {#ex_script}

	Meget uenig	Overvejende uenig	Overvejende enig	Meget enig	Ved ikke
Jeg har en særlig person, som jeg kan gå til, hvis jeg har behov for det	☐	☐	☐	☐	☐
Jeg har en særlig person, som jeg kan dele sorger og glæder med	☐	☐	☐	☐	☐
Min familie prøver virkelig på at hjælpe mig	☐	☐	☐	☐	☐
Jeg får den følelsesmæssige hjælp og støtte, som jeg har behov for fra min familie	☐	☐	☐	☐	☐
Jeg har en særlig person, som er en virkelig stor støtte for mig	☐	☐	☐	☐	☐
Mine venner prøver virkelig på at hjælpe mig	☐	☐	☐	☐	☐
Jeg kan regne med mine venner, når noget går galt	☐	☐	☐	☐	☐
Jeg kan tale med min familie om mine problemer	☐	☐	☐	☐	☐
Jeg har venner, som jeg kan dele sorger og glæder med	☐	☐	☐	☐	☐
Der er en særlig person i mit liv, som	☐	☐	☐	☐	☐

bekymrer sig om mine følelser					
Min familie er villig til at hjælpe mig med at træffe beslutninger	☐	☐	☐	☐	☐
Jeg kan tale med mine venner om mine problemer	☐	☐	☐	☐	☐

[END PAGE – q17_page]

[q18 – CATEGORICAL – single – Must answer]

Hvordan er dit helbred generelt? Vil du sige, at det er:

- ☐ (1) Meget godt
- ☐ (2) Godt
- ☐ (3) Rimeligt
- ☐ (4) Dårligt
- ☐ (5) Meget dårligt
- ☐ (6) Ved ikke

[q19 – CATEGORICAL – single – Must answer]

Føler du dig stresset?

- ☐ (1) Nej, overhovedet ikke
- ☐ (2) Ja, lidt
- ☐ (3) Ja, meget
- ☐ (4) Ved ikke

[q20 – CATEGORICAL – single – Must answer]

Hvor ofte er du stresset?

- ☐ (1) Aldrig/næsten aldrig
- ☐ (2) Månedligt
- ☐ (3) Ugentligt
- ☐ (4) Dagligt
- ☐ (5) Ved ikke

[q21 – CATEGORICAL – single – Must answer]

Er der sket vigtige ting for dig, en meget nær ven eller tæt familie inden for de sidste 12 måneder, som har fået den periode til at være betydeligt vanskeligere for dig end et typisk år?

- ☐ (1) Ja
- ☐ (2) Nej

[q22 – CATEGORICAL – single – Must answer]

I hvilken grad oplever du i øjeblikket økonomiske problemer?

- ☐ (1) I meget høj grad
- ☐ (2) I høj grad
- ☐ (3) I nogen grad
- ☐ (4) I mindre grad
- ☐ (5) Slet ikke
- ☐ (6) Ved ikke

[q23_ – CATEGORICAL – multiple – Must answer]

Hvem bor du sammen med i øjeblikket? <i>Mulighed for flere svar.</i>

- ☐ (1) Mine forældre (eller bedsteforældre, onkler, tanter eller lign.)
- ☐ (2) Partner
- ☐ (3) Egne/partners barn/børn
- ☐ (4) Ven(ner)/veninde(r)

- (_5) Andre
- (_6) Jeg bor alene

[If not(q23_Response.Value.ContainsAny(_1)) Then]

[q24 – CATEGORICAL – single – Must answer]

Hvordan bor du i øjeblikket?

- o (_1) På et kollegium
- o (_2) I en studiebolig/ungdomsbolig
- o (_3) Jeg lejer et værelse
- o (_4) Jeg lejer en lejlighed
- o (_5) I egen andelsbolig
- o (_6) I egen ejerbolig
- o (_9) Andet

[End If]

[W1_New1 – CATEGORICAL – single – Must answer]

Forventer du at skulle flytte, eller er du flyttet i forbindelse med studiestart?

- o (_1) Jeg er flyttet
- o (_2) Jeg forventer at flytte
- o (_3) Jeg skal ikke flytte
- o (_4) Ved ikke

[q25 – CATEGORICAL – single – Must answer]

I hvilken grad er du bekymret over din boligsituation?

- o (_1) I meget høj grad
- o (_2) I høj grad
- o (_3) I mindre grad
- o (_4) Slet ikke
- o (_5) Ved ikke

[q26 – CATEGORICAL – single – Must answer]

Hvad er den gennemsnitlige transporttid, du forventer fra din bolig til dit studie, når du starter på studiet?

- o (_1) Notér antal minutter:
- o (_2) Ved ikke

[outro – CATEGORICAL – single – Must answer]

Tak for din besvarelse! Vi håber, at du havde en god oplevelse med at besvare spørgsmålene. Som nævnt i invitationsbrevet vil vi gerne følge dig i din første tid på studiet – vi er nemlig nysgerrige på dine erfaringer og oplevelser. Må vi have lov til at kontakte dig igen?

- o (_1) Ja, det må I gerne
- o (_2) Nej, jeg ønsker ikke at blive kontaktet igen

1.1.1..2 <EndOfSurveyBM:>

<Goto ENDQ>

1.1.1..3 <TerminateScriptBM:>

<Goto ENDQ>

1.1.1..4 <QuotaFailScriptBM:>

<Goto ENDQ>

1.1.1..5 <Default_ErrorHandler:>

<Goto ENDQ>

1.1.1..6 <ENDQ:>

Tabel .1: Balancetabel

	Stikprøve med spørgeskemadata		Stikprøve		Population	
	Middel	sd.	Middel	sd.	Middel	sd.
Mand	0,40	0,49	0,46	0,49	0,48	0,49
Alder	21,53	0,02	21,62	0,02	21,84	0,01
Dansk oprindelse	0,90	0,32	0,88	0,33	0,87	0,35
Mor har en videregående uddannelse	0,21	0,41	0,21	0,41	0,20	0,40
Far har en videregående uddannelse	0,29	0,45	0,29	0,45	0,28	0,45
Mors log-indkomst ved det 15. leveår	12,81	0,00	12,81	0,00	12,78	0,00
Fars log-indkomst ved det 15. leveår	13,05	0,01	13,06	0,00	13,02	0,00
1. prioritet	0,84	0,00	0,83	0,78	0,83	0,80
Søgt kvote 2	0,31	0,46	0,30	0,46	0,30	0,46
Optaget i kvote 2	0,14	0,35	0,16	0,36	0,14	0,34
Førsteårsfrafald	0,16	0,37	0,18	0,38	0,19	0,39
Antal observationer	26.011		51.023		69.769	

Note: .

Kilde: Egne beregninger baseret på registerdata fra Danmarks Statistik

Tabellen ovenfor viser, hvor repræsentativ stikprøven af respondenter som har besvaret spørgeskemaet er sammenlignet med den fulde population.

Appendiks B: Empirisk strategi

Det fuzzy regression diskontinuitet design

I dette speciale anvendes det skarpe regression diskontinuitet design, da data udelukkende indeholder information om optagne studerende på en universitetsuddannelse. Dette medfører, at data ikke indeholder informationer om ansøgere, som ikke eksempelvis har valgt at takke nej til en studieplads selvom vedkommende havde et karaktergennemsnit, som opfyldte adgangskvotienten på den ansøgte uddannelse. Derudover er der i datarestriktionen fjernet ansøger, som har sikret deres optagelse til en studieplads gennem en standbyplads. Denne restriktion er foretaget for at sikre et homogent datasæt.

Appendiks C: Resultater

Tabel .2: Effekten af optagelseskvoten på førsteårs frafaldssandsynlighed for universitetsbacheloror for det simple RD design med forskellige båndbredde, 2016-2018

	(1)	(2)	(3)	(4)
RD koeficient	0,055*** (0,014)	0.050*** (0,017)	0,052* (0,028)	0,078 (0,050)
Båndbredde, h	1,347	1	0,5	0,25
Demografiske kontroller	Nej	Nej	Nej	Nej
Selvrapporert afklarethed	Nej	Nej	Nej	Nej
Fixed effects	Nej	Nej	Nej	Nej
Antal effektive observationer	27.124	20.468	10.868	6.534
Til venstre for cut-off	3.773	2.739	1.242	582
Til højre for cut-off	23.351	17.729	9.626	5.952

Note: Robuste standardfejl er angivet i parentes. * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$.

Tabel .3: Effekten af optagelseskvoten på førsteårs frafaldssandsynlighed for universitetsbacheloror for RD designet med fixed effects for forskellige båndbredde, 2016-2018

	(1)	(2)	(3)	(4)
RD koeficient	0,048*** (0,011)	0.043*** (0,014)	0,047** (0,028)	0,070* (0,039)
Båndbredde, h	1,344	1	0,5	0,25
Demografiske kontroller	Nej	Nej	Nej	Nej
Selvrapporert afklarethed	Nej	Nej	Nej	Nej
Fixed effects	Ja	Ja	Ja	Ja
Antal effektive observationer	27.124	20.468	10.868	6.534
Til venstre for cut-off	3.773	2.739	1.242	582
Til højre for cut-off	23.351	17.729	9.626	5.952

Note: Robuste standardfejl er angivet i parentes. * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$.

Tabel .4: Effekten af optagelseskvoten på førsteårs frafaldssandsynlighed for universitetsbachelorer for RD designet med fixed effects og demografiske kontroller for forskellige båndbredde, 2016-2018

	(1)	(2)	(3)	(4)
RD koeficient	0,043*** (0,011)	0.038*** (0,014)	0,040* (0,022)	0,056 (0,039)
Båndbredde, h	1,344	1	0,5	0,25
Demografiske kontroller	Ja	Ja	Ja	Ja
Selvrapporteret afklarethed	Nej	Nej	Nej	Nej
Fixed effects	Ja	Ja	Ja	Ja
Antal effektive observationer	27.124	20.468	10.868	6.534
Til venstre for cut-off	3.773	2.739	1.242	582
Til højre for cut-off	23.351	17.729	9.626	5.952

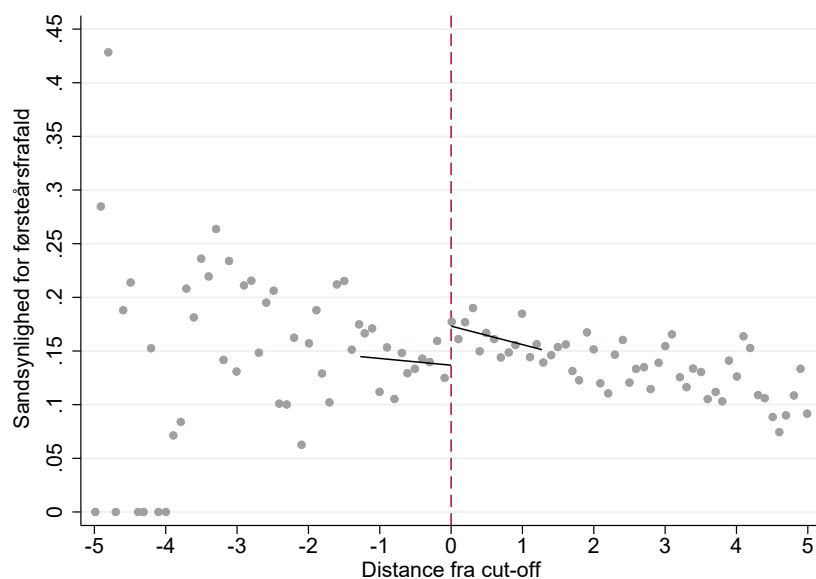
Note: Robuste standardfejl er angivet i parentes. * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$.

Tabel .5: Effekten af optagelseskvoten på førsteårs frafaldssandsynlighed for universitetsbachelorer for RD designet med fixed effects, demografiske kontroller og selvrapporteret afklarethed for forskellige båndbredde, 2016-2018

	(1)	(2)	(3)	(4)
RD koeficient	0,042*** (0,015)	0.033* (0,018)	0,012 (0,030)	0,018 (0,078)
Båndbredde, h	1,369	1	0,5	0,25
Demografiske kontroller	Ja	Ja	Ja	Ja
Selvrapporteret afklarethed	Ja	Ja	Ja	Ja
Fixed effects	Ja	Ja	Ja	Ja
Antal effektive observationer	13.367	10.005	5.295	3.882
Til venstre for cut-off	1.839	1.351	627	320
Til højre for cut-off	11.528	8.654	4.668	3.562

Note: Robuste standardfejl er angivet i parentes. * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$.

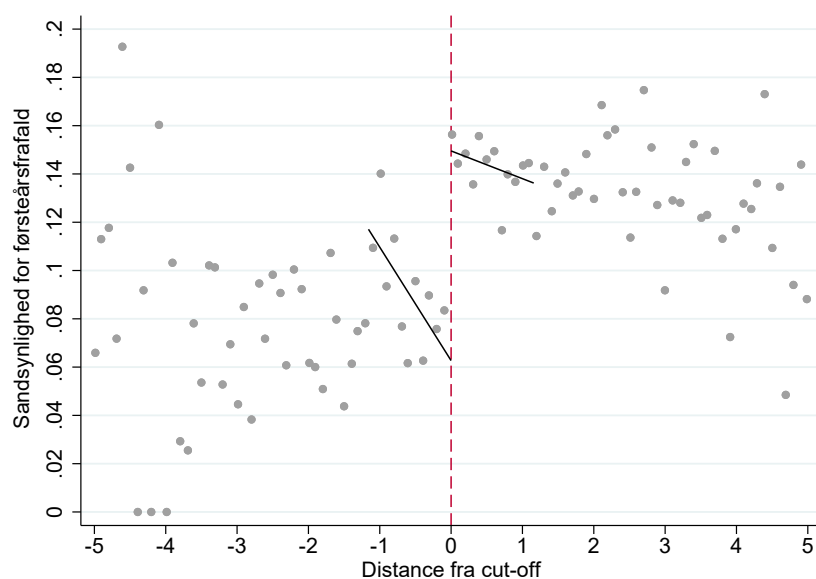
Figur .1: Optagelseskvoters effekt på førsteårs frafaldssandsynligheden for universitetsbachelorer med passiv uden motiveret ansøgning, 2016-2018



Note: Figuren viser en simpel sammenhæng mellem at frafalde påbegyndt uddannelse inden for et studieår for studerende optaget i kvote 2 (venstre) og kvote 1 (højre). Der er anvendt et fixed effects RD design. Den optimale MSE båndbredde er 1.273

Source: DST

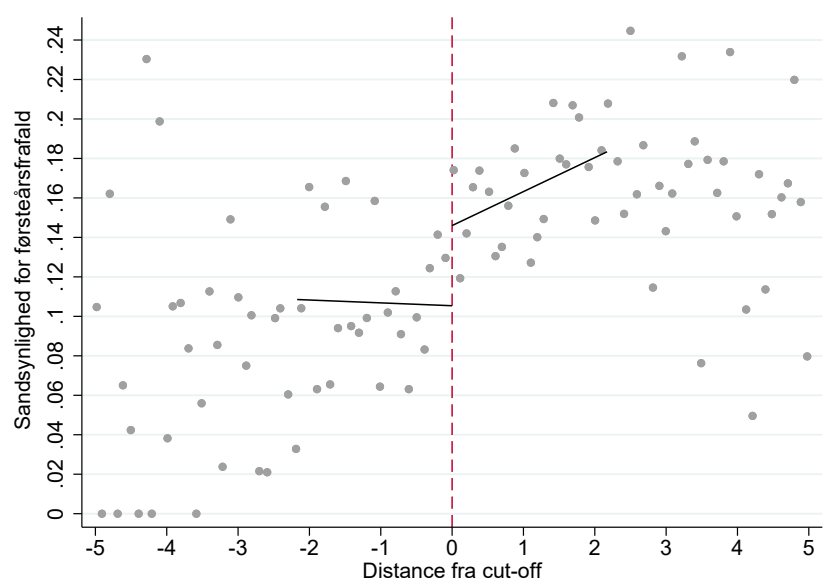
Figur .2: Optagelseskvoters effekt på førsteårs frafaldssandsynligheden for universitetsbachelorere med kvote 2 med passiv med motiveret ansøgning, 2016-2018



Note: Figuren viser en simpel sammenhæng mellem at frafalde påbegyndt uddannelse inden for et studieår for studerende optaget i kvote 2 (venstre) og kvote 1 (højre). Der er anvendt et fixed effects RD design. Den optimale MSE båndbredde er 1.158

Source: DST

Figur .3: Optagelseskvoters effekt på førsteårs frafaldssandsynligheden for universitetsbachelorer med kvote 2 med fremmøde, 2016-2018



Note: Figuren viser en simpel sammenhæng mellem at frafalde påbegyndt uddannelse inden for et studieår for studerende optaget i kvote 2 (venstre) og kvote 1 (højre). Der er anvendt et fixed effects RD design. Den optimale MSE båndbredde er 2.172

Source: DST