



Kandidatafhandling

Anne Kristine Tiedemann Skovbon

Social skævhed i frafaldssandsynligheden på universitetet

En økonometrisk analyse af universiteternes bacheloruddannelser

Afleveret den 21. december 2020

Vejleder: Mette Ejrnæs

Antal tegn: 129.891

Abstract

Recent Danish literature suggest that the social mobility has been declining during the last decade. In terms of social mobility, ensuring that individuals have equal opportunities in educational is a key instrument. However, simply focusing on the enrollment in higher education is not sufficient to ensure equality in opportunity. The purpose of this thesis is to examine social inequality in university dropout and whether the dropout's dependence on parents' educational attainment changes throughout the time of study.

For this purpose, detailed administrative data from Statistics Denmark is employed. The analysis is based on a sample of students enrolled in 2016 and 2017 in a bachelor's programme at one of the eight Danish universities. A logit model controlling for relevant background characteristics and unobserved heterogeneity across the different studies, is used for estimating the probability of dropping out. The probability is estimated at six different time periods during the first two years of study. The timing of the dropout is not treated as continuous in time since there typically exists a lag in when the dropout occurs and when the dropout is administratively registered. Due to this inconvenience of the dropout date registered by Statistics Denmark, a duration model will not be used. Instead, the two first years of studies are divided into six different time periods defined by the ending of semesters. Using maximum-likelihood estimation, the probability of dropping out is estimated for the six different time periods. The main variable of interest is the parents' highest level of educational attainment which is defined by the parent with the highest educational attainment.

The maximum-likelihood estimations reveal that parents' educational background only tends to be of significant meaning for dropout during the first semester. Students whose parents have compulsory schooling as their highest level of educational attainment, have 30% higher probability of dropping out during the first semester compared to students whose parents have completed a long-cycle education. Students whose parents have completed a long-cycle education have significantly lower probability of dropping out during the first semester of their studies compared to any other student whose parents have lower educational attainment. The probability of dropping out does not depend on parents' educational attainment for any of the time periods not containing the first semester. It appears that students' social background only matters for the early dropout.

The findings of the econometric analysis suggest that more political focus on the social inequality in university dropout is needed. The findings stress that students meet academia with different social backgrounds, and thus different conditions for completing their studies. For a more efficient

political intervention, political effort to reduce the social gap in university dropout should be concentrated around the first semester. A possible political intervention which has shown effective according to the existing literature, is academic preparatory course targeted students whose parents have low educational attainment.

Jeg vil gerne give en stor tak til Danmark Evalueringsinstitut, EVA, for samarbejdet om min specialeafhandling. Tak for god og motiverende sparring samt at have stillet data til rådighed. Endeligt tak til min vejleder Mette Ejrnæs for kyndig og konstruktiv vejledning.

Social skævhed i frafaldssandsynligheden på universitetet

EN ØKONOMETRISK ANALYSE AF UNIVERSITETERNES BACHELORUDDANNELSER

Anne Kristine Tiedemann Skovbon

Københavns Universitet

Den 21. december 2020

Indhold

1	Indledning	1
2	Baggrund	3
2.1	Den danske universitetssektor	3
2.2	Optagelse på videregående uddannelser	3
3	Teoretisk framework	5
3.1	Humankapitalteorien	5
3.2	Teorien om relativ risikoaversion	6
3.3	Tintos frafaldsmodel for videregående uddannelse	7
3.4	Teoretiske hypoteser	7
4	Litteratur gennemgang	9
4.1	Årsager til frafald	9
4.2	Betydningen af forældrebaggrund for frafald	10
4.3	Bidrag til den eksisterende litteratur	11
5	Data	12
5.1	Populationsafgrænsning	12
5.2	Anvendte empiriske variable	12
5.2.1	Den afhængige variabel	12
5.2.2	Kontrolvariable	14
6	Deskriptiv analyse af data	17
7	Økonometrisk metode	19
7.1	Logitmodellen	19
7.2	Den økonometriske model	22

7.2.1	Endogenitetsudfordringer	22
7.2.2	Clustering og fixed-effects	23
8	Resultater	24
8.1	Hovedresultater	24
8.2	Delanalyse: forskelle på tværs af hovedområder	30
8.3	Delanalyse: de studerende som falder fra	34
9	Robusthedsanalyse	37
9.1	Indkomst som mål for forældrebaggrund	37
9.2	Fixed-effects på alternativt uddannelsesudbud-niveau	38
9.3	Kontrol for studerendes bopæl inden studiestart	39
10	Diskussion	41
10.1	Data- og metodiske overvejelser	41
10.2	Validitet af hovedresultater	42
10.3	Politiske anbefalinger	43
10.4	Videre forskning	44
11	Konklusion	45
12	Litteraturliste	47
A	Appendiks	51
A.1	Regressionstabeller	51
A.2	Diverse	56

Figurer

1	Perioder frafaldssandsynligheden estimeres	13
2	Frafald fordelt på optagne studerendes forældres højeste fuldførte uddannelsesniveau	18
3	Prædikterede marginale sandsynligheder med 95% konfidensinterval - grundskole som referencekategori	26
4	Andetsårsfrafald fordelt på optagne studerendes forældres højeste uddannelsesniveau, opdelt på hovedområder	31
5	Prædikterede marginale sandsynligheder uden referencekategori med 95% konfidensinterval - opdelt på universiteternes hovedområder	32
6	Fordeling i uddannelsestype påbegyndt eller færdiggjort efter frafald, fordelt på forældres højeste fuldførte uddannelsesniveau - ultimo 2018	35
7	Fordeling i de frafaldne studerendes tilknytning til arbejdsmarkedet, fordelt på forældres højeste fuldførte uddannelsesniveau - ultimo 2018	36
8	Prædikterede marginale sandsynligheder uden referencekategori med 95% konfidensinterval	56
9	Histogram for frafaldsdatoen	57

Tabeller

1	Frafald for de seks perioder, optag 2016 og 2017	14
2	Beskrivende statistik for datagrundlaget $N=42.675$	17
3	Optag og fuldførelse fordelt på optagne studerendes forældres højeste fuldførte uddannelsesniveau	18
4	Kvadreret Wald test af forældres højeste fuldførte uddannelsesniveau	28
5	Prædikterede marginale sandsynligheder med diverse uddannelsesniveauer som referencekategori for forældres højeste fuldførte uddannelsesniveau, 1. semester frafaldssandsynlighed	29
6	Fordelingen i forældres højeste fuldførte uddannelsesniveau blandt optag fordelt på hovedområder	30
7	Kvadreret Wald test af forældres højeste fuldførte uddannelsesniveau for hovedområderne	33
8	Andel blandt frafald som påbegynder en ny videregående uddannelse der ikke afbrydes, fordelt på forældres højeste fuldførte uddannelsesniveau - ultimo 2018	34
9	Kvadreret Wald test for forældres højeste fuldførte uddannelsesniveau - model med fixed-effects på hovedområde-institution	39
10	Regressionstabel, gennemsnitlige marginale sandsynligheder	52
11	Regressionstabel - hovedområde subsamples, gennemsnitlige marginale sandsynligheder	53
12	Regressionstabel - tilflytter-variabel og fixed-effects på hovedområde-institution, gennemsnitlige marginale sandsynligheder	54
13	Regressionstabel - forældres indkomstsniveau, gennemsnitlige marginale sandsynligheder	55
14	Fordelingen i forældres højeste fuldførte uddannelsesniveau blandt optag fordelt på hovedområder	57

15	Kvadreret Wald test af mors indkomstniveau (log-transformeret)	57
16	Kvadreret Wald test af fars indkomstniveau (log-transformeret)	58

1 Introduktion

Den danske velfærdsstat er bygget op omkring idealet om lige muligheder for alle. Det skal være individets evner og ambitioner, som afgør, hvor vedkommen ender i samfundet. Her ses uddannelse ofte som et af de vigtigste redskaber. Det er svært at overdrive betydningen af uddannelse - både for individets livsforløb og for samfundsøkonomien. Individet gavner både af indkomstfortjenesten og jobsikkerheden ved højere uddannelsesniveau samt afsmittende effekter såsom sundere livsstil, generelt bedre helbred, mindre kriminalitet og større samfundsengagement [Lochner, 2011]. Et øget uddannelsesniveau i befolkningen gavner samfundsøkonomien gennem bl.a. øget skatteprovenu, højere produktivitet i samfundet og lavere udgifter til sundhedsvæsenet og politi [Sørensen et al., 2013].

Flere nyere danske undersøgelser har dokumenteret, at den sociale mobilitet i Danmark er blevet forringet de seneste 10-15 år [Pedersen and Caspersen, 2020, Karlson and Landersø, 2020, DEA, 2019]. På trods af gratis adgang til videregående uddannelse og mange årtiers politiske indsatser, eksisterer der stadig en stærk sammenhæng mellem, hvor individet starter og ender i samfundet. De nuværende politiske indsatser er ikke tilstrækkelige til at indfri velfærdsstatens ideal. Her er uddannelse en af de vigtigste katalysatorer, og uddannelsessystemet har sin største udfordring med at mindske betydningen af individets sociale baggrund i universitetssektoren. Et relevant fokus er derfor, hvor der kan rettes ind for at mindske den sociale arvs tag i universiteterne.

Frafaldet på universiteterne har længe været et politisk fokusområde. Når den studerende afbryder en uddannelse, vedkommende ellers var kvalificeret til at fuldføre, er det et nederlag for individet, som går glip af uddannelsens positive effekter for livsforløbet, for samfundet som går glip af individets humankapital og for uddannelsesinstitutionerne, hvor frafaldet er forbundet med et finansielt tab. Fravalddet er af stor samfundsmæssig bekymring, hvilket forstærkes yderligere ved den sociale skævhed i frafaldet. Selv når der kontrolleres for forskelle i de studerendes faglige evner, ses der en social slagside i universiteternes frafald [DEA, 2020, Smith et al., 2020, Videnskabsministeriet, 2008, Ulriksen et al., 2015]. De studerendes sociale baggrund betyder, at de møder universitetsverden med forskellige forudsætninger, hvilket kan have betydning for, hvordan de begår sig i den akademiske verden, og hvorvidt de lykkes med at fuldføre uddannelsen.

I nærværende afhandling undersøges den sociale skævhed i frafaldet på universiteternes bacheloruddannelser med udgangspunkt i problemformuleringen: *Hvordan ændres betydningen af forældrebaggrund for frafaldssandsynligheden i løbet af studietiden?* Problemformuleringen forsøges besvaret ud fra maximum-likelihood estimation af en logitmodel, hvis binære afhængige variabel beskriver den studerendes sandsynlighed for frafald. Fravalddssandsynligheden estimeres for seks forskellige perioder. De to første studieår inddrages i disse seks perioder, som afgrænses ved semesterskift. Denne tilgang tager højde for, at den eksakte dato for frafald ikke er administrativt registreret, men typisk registreres med en tidsmæssig forskydning. Der kontrolleres for en række relevante baggrundskarakteristika samt uobserveret heterogenitet på tværs af uddannelsesudbud gennem en fixed-effects tilgang.

Analysen bidrager til den nyere danske frafaldslitteratur ved sit fokus på, om betydningen af forældrebaggrund for frafald *ændres* i løbet af studietiden. Ved politisk intervention vil det have et stort potentiale

at vide, *hvornår* det vil være mest efficient at sætte ressourcer ind. Et yderligere bidrag til den eksisterende danske frafaldslitteratur er, at analysens datagrundlag ikke er afgrænset til enkelte universiteter eller hovedområder, som det er tilfældet for mange af de eksisterende danske frafaldsstudier med fokus på social skævhed. Slutteligt er nærværende afhandling med til at understrege vigtigheden i, at der sættes ind mod den sociale skævhed i frafaldet på universiteterne. På trods af stort politisk fokus på at mindske de videregående uddannelsers frafald, er der et meget beskedent fokus på frafaldsindsatser, som er målrettet studerende fra ikke-uddannelsesvante hjem [DEA, 2020].

2 Baggrund

Dette afsnit introducerer relevante aspekter af den danske universitetssektor og optagelsessystemet for de videregående uddannelser.

2.1 Den danske universitetssektor

I Danmark er der otte universiteter, som er omfattet af universitetsloven: Københavns Universitet (KU), Aarhus Universitet (AU), Aalborg Universitet (AAU), Syddansk Universitet (SDU), Roskilde Universitet (RUC), Copenhagen Business School (CBS), Danmarks Tekniske Universitet (DTU) og IT-Universitetet i København (ITU). Forskellen mellem universiteterne afspejler sig både i deres størrelse samt faglige, regionale og historiske profil. På de danske universiteter udbydes der akademiske lange videregående uddannelser, som er normeret til 5 år. Bacheloruddannelserne er 3-årige uddannelser, der kan overbygges med en 2-årige kandidatuddannelse, hvilket kan efterfølges af en af ph.d. [Danske Universiteter, 2020]. Bacheloruddannelsernes opbygning varierer mellem blok- og semesterstruktur. På de fleste uddannelser er studieåret opdelt i semestre med to semestre på et studieår. Ved blokstruktur er studieåret opdelt i fire blokke, hvor nogle kurser løber over to blokke svarende til et semester. Eksaminer, offentliggørelse af eksamensresultater og kursustilmeldinger sker ved blok- og semesterskift.

I universitetssektoren opereres der med hovedområder, hvilke betegner de overordnede fagfællesskaber, der ofte også er inddelingen for universiteternes fakulteter. Typisk indeles der i fem hovedområder: humaniora, naturvidenskab, samfundsvidenskab, sundhedsvidenskab og teknisk videnskab.

I Danmark er det gratis at tage en videregående uddannelse, og siden 1970 har studerende kunne modtage statens uddannelsesstøtte (SU) under deres studietid. SU-ordningen blev indført for i højere grad at sikre, at alle individer har økonomisk og socialt lige muligheder for uddannelse.

Der har været et svagt fald i universiteternes optag siden 2013, men set over en tiårige periode er optaget steget. I 2009 blev 22.353 studerende optaget på en bacheloruddannelse, hvilket er steget med 34 pct. til 29.876 optagne i 2019. Det stigende optag har bl.a. medført stigende adgangskvotienter. Andelen af uddannelser med adgangskvotient over 9,0 er steget fra at udgøre 6 pct. af alle uddannelser i 2009 til knap 17 pct. i 2019. Siden år 2000 har der været flere kvindelige end mandlige studerende på universiteterne. Kønsfordelingen blandt de universitetsstuderende har siden 2013 ligget stabilt omkring 53 pct. kvinder mod 47 pct. mænd. De senest ti år har gennemsnitsalderen på studerende, som påbegynder en bacheloruddannelse ligget stabilt omkring 21 år [Danske Universiteter, 2020].

2.2 Optagelse på videregående uddannelser

Ansøgning og optagelse på alle videregående uddannelser under Uddannelses- og Forskningsministeriet bliver koordineret ved Den Koordinerede Tilmelding (KOT). Hver ansøger kan søge om optagelse på op til 8 forskellige uddannelser, som rangeres efter prioritet. Ansøgninger opdeles i kvote 1 og kvote 2. Hvis der er flere ansøgere end der er studiepladser, fordeles kvote 1 studiepladser efter de adgangsgivende karaktergennemsnit, hvor ansøgeren med det højeste karaktergennemsnit først tildes en studieplads. Karaktergennemsnittet for den sidste studerende, som bliver tildelt en studieplads, udgør årets adgangskvo-

tient. I kvote 2 vurderes ansøgerne på andre adgangsgrundlag end karaktergennemsnittet såsom faglige forudsætninger og motivation for ansøgning til uddannelsen. Er der færre ansøgere end der er studiepladser, bliver alle kvalificerede ansøgere optaget og det registreres ikke, om de er optaget på kvote 1 eller kvote 2.

3 Teoretisk framework

Den teoretiske ramme for afhandlingens problemformulering opstilles i dette afsnit. Der inddrages økonomisk og sociologisk teori for at opnå en forståelse af, hvorfor individer vælger at påbegynde og afbryde en videregående uddannelse. Slutteligt opstilles de teoretiske hypoteser for afhandlingens problemformulering.

3.1 Humankapitalteorien

Humankapitalteorien blev udviklet af G. Becker i 1960'erne [Becker, 1964] og bygger på antagelsen om, at individets uddannelsesvalg er baseret på rationelle overvejelser om, hvorvidt en investering i uddannelse kan betale sig. Uddannelse betragtes som en privat investering, hvor den fremtidige øget indkomst er afkastet, hvilket individet ønsker at maksimere. Afkastet afvejes mod uddannelsens omkostninger, som kan inddrages i finansielle og ikke-finansielle. I det danske uddannelsessystem er de direkte finansielle uddannelsesomkostninger minimeret grundet gratis uddannelse gennem hele uddannelsessystemet. Den indirekte finansielle omkostning er alternativomkostningen af den tid, som individet har investeret i uddannelsen dvs. den tabte indkomst som uafaglært i tiden som studerende. De ikke-finansielle omkostninger er den studerendes indsats og anstrengelser i at fuldføre uddannelsen og eventuel misfornøjelse ved at studere. Individet vælger at påbegynde en videregående uddannelse på baggrund af en begrundet vurdering af, om det forventede afkast af uddannelsen overstiger de forventede finansielle og ikke-finansielle omkostninger [Mattsson and Munk, 2008].

I humankapitalteorien antages uddannelsesbeslutning at være statisk. I en model præsenteret af Aina et al. [Aina et al., 2018] inddrages uddannelsesbeslutningen som en dynamisk proces, hvor påbegyndelse, fuldførelse og frafald er resultater af løbende opdaterede forventninger til det fremtidige afkast og omkostninger ved uddannelsen. I den dynamiske model antages det, at den studerende gennem hele studietiden gennemgår en gradvis læringsproces, som den studerende løbende reviderer sin uddannelsesbeslutning på baggrund af. Den studerende lærer gradvist om egne evner i relation til den valgte uddannelse gennem karaktermæssig tilbagemelding, om fornøjelse og misfornøjelse ved at studere og om den forventede fremtidige indkomst forbundet med den valgte uddannelse. Det fører til, at den studerende på baggrund af rationelle overvejelser løbende vurderer, hvorvidt investeringen i uddannelsen kan betale sig. Når de finansielle og ikke-finansielle omkostninger overstiger det forventede afkast, vil den studerende beslutte at afbryde uddannelsen. Denne dynamiske model tager højde for, at den studerende inden studiestart ikke har perfekt information om kvaliteten af uddannelsen, sin oprigtige interesse i uddannelsens faglige indhold, hvorvidt ens evner imødekommer uddannelsens faglige krav samt de påkrævede anstrengelser i at studere.

I flere frafaldsstudier præsenteres der økonomiske beslutningsmodeller, hvor uddannelsesbeslutningen løbende opdateres på baggrund af en gradvis læringsproces i løbet af studietiden. I forskningsartiklen af Stinebrickner & Stinebrickner [Stinebrickner and Stinebrickner, 2012] undersøges betydningen af de studerendes læringsproces på baggrund af karaktergivning på 1. studieår på universitetet. Ligeledes opstiller Arcidiacono et al. [Arcidiacono et al., 2016] en dynamisk strukturel beslutningsmodel, hvor individerne har imperfekt information om egne akademiske evner på universitetet. De inddrager i modellen

individets læring af egne evner gennem karakterer, hvorved individet kan vælge at afbryde uddannelsen, som resultat af den nye information om egne akademiske evner.

3.2 Teorien om relativ risikoaversion

Sociologiens teori om relativ risikoaversion (RRA) tager afsæt i humankapitalteorien. RRA teorien blev oprindeligt formuleret af R. Boudon i 1970'erne [Boudon, 1974], og blev senere videreudviklet af R. Breen og J. Goldthorpe [Breen and Goldthorpe, 1997].

Ifølge RRA teorien træffes uddannelsesvalg ud fra en rationel vurdering af forventet afkast, omkostninger og sandsynlighed for succes ved forskellige uddannelsesalternativer [Mattsson and Munk, 2008]. RRA teorien har en mere social dimension end humankapitalteorien, hvor det antages, at alle individer handler ud fra samme rationalitet. I RRA tillades en usikkerhed om sandsynligheden for succesfuld fuldførelse af en specifik uddannelse. Derved kan individer med forskellige sociale baggrunde have forskellige rationelle vurderinger af samme uddannelsesalternativ. Individet har en stærk præference for at minimere risikoen for nedadgående mobilitet dvs. ende med et lavere uddannelsesniveau end forældrene. Individet tilstræber at opnå *mindst* samme uddannelsesniveau eller social position som forældrene uanset om individets evner berettiger til et højere uddannelsesniveau. Børn fra hjem uden tradition for at tage en universitetsuddannelse vil opfatte det som mere risikofyldt at påbegynde en universitetsuddannelse end børn fra akademikerhjem vil, da der er risiko for ikke at gennemføre uddannelsen [Holm and Jæger, 2008].

RRA teorien beskriver denne tendens ved et *knæpunkt*. Så længe barnet ikke har opnået et tilsvarende uddannelsesniveau som forældrene, vil barnet træffe uddannelsesvalg, der maksimerer sandsynligheden for at opnå samme uddannelsesniveau. Når barnet har opnået samme uddannelsesniveau som forældrene vil incitamentet til at uddanne sig videre aftage. Humankapitalteorien antager en lineær, proportional sammenhæng mellem investeringen i uddannelse og afkastet heraf. Som modstykker hertil antager RRA teorien, at marginalnyttens ved yderligere uddannelsesinvesteringer aftager, når barnet opnår samme uddannelsesniveau som forældrene og dermed eliminerer risikoen for nedadgående mobilitet [Mattsson and Munk, 2008].

I teorien om RRA opdeles effekten af social baggrund på opnået uddannelsesniveau i primære- og sekundære effekter [Thomsen and Andrade, 2016]. Den primære effekt er sammenhængen mellem et barns sociale baggrund og dets kognitive og ikke-kognitive evner. Der eksisterer en tendens til, at børn fra lavere sociale klasser har lavere kognitive og ikke-kognitive evner sammenlignet med børn fra højere sociale klasser. Den sekundære effekt er den uddannelsesmæssige ulighed, som vedbliver efter der kontrolleres for klasseforskelle i evner. Effekten betyder, at børn fra højere sociale klasser vil opnå højere uddannelsesniveauer end børn med samme evner fra lavere sociale klasser. Derved kan forskelle i ambitioner og præferencer for uddannelsesvalg, fremsættes som forklaring på, at børn af samme evner med forskellige sociale baggrunde opnår forskellige uddannelsesniveauer [Mattsson and Munk, 2008].

3.3 Tintos frafaldsmodel for videregående uddannelse

Uddannelsessociologen Vincent Tinto har udviklet en model for forståelse af hvilke faktorer, der spiller ind i processen frem mod den studerendes beslutning om at forlade en videregående uddannelse. Ifølge Tinto sker det frivillige frafald som følge af et ikke-succesfuldt møde mellem den studerende og uddannelsen. Helt centralt er, at frafaldet sker i samspillet mellem den studerende og uddannelsen.

Analytisk skelner modellen mellem den studerendes individuelle baggrundsfaktorer - intelligens, motivation og forældrebaggrund - og den sociale og faglige integration på uddannelsen. Frafaldet sker i en proces mellem de individuelle baggrundsfaktorer og mødet med uddannelsen og det akademiske systems normer. Mødet med uddannelsen og det akademiske systems normer resulterer i varierende grader af social og faglig integration. En vellykket integration fremmer den studerendes tilknytning og engagement i uddannelsen, hvilket øger vedholdenheden på uddannelsen. Tinto fremhæver bl.a. følelsen af at høre til som centralt for den faglige og sociale integration. Følelsen af at høre til refererer til, i hvilken grad den studerende ser sig selv som en del af et fællesskab med andre på uddannelsen, hvilket skal være både fagligt og socialt funderet, for at integrationen er succesfuld [Tinto, 1975, Ulriksen et al., 2015].

I Tintos model lægges der vægt på, at integrationen afhænger af individets individuelle baggrundsfaktorer. Forældrebaggrund kan påvirke graden af den studerendes sociale- og faglige integration gennem flere mekanismer. Forældre som har haft et længere uddannelsesforløb vil, alt andet lige, have mere kendskab og erfaring med uddannelse. Forældrenes kendskab og erfaring kan være i form af konkret information om afkastet af uddannelse, og hvad det kræver at fuldføre en uddannelse, eller af mere følelsesmæssig karakter såsom følelsen af at høre til på en uddannelse. På den baggrund kan forældrene bedre navigere og vejlede deres børn igennem uddannelsessystemet [Karlson, 2014, Danmarks Evalueringsinstitut, 2020]. Foruden forældrenes kendskab og erfaringer med uddannelsessystemet kan deres høje uddannelsesniveau også påvirke barnets uddannelsesvalg gennem den tidlige socialisering i familien. Forældre med lange uddannelser kultiverer i højere grad uddannelsesrelevante evner hos deres børn gennem familiens opdragelsesmønstre og hverdagsstruktur. Det kan være evner såsom at have et større ordforråd, være mere tålmodig og at kunne tilpasse sig sociale sammenhænge [Thomsen and Bom, 2019, Danmarks Evalueringsinstitut, 2020]. Disse uddannelsesrelevante evner individet har tilegnet sig gennem opvæksten samt forældrenes kendskab til universitetsverden, kan betyde, at studerende med akademikerforældre i højere grad oplever et succesfuldt møde med universitetsverden, mens studerende uden akademikerforældre kan have sværere ved at afkode det akademiske systems normer og derfor i mindre grad oplever at høre til på universitet [Smith et al., 2020].

3.4 Teoretiske hypoteser

På baggrund af ovenstående teoretiske betragtninger opstilles følgende teoretiske hypoteser:

1. Individuer med akademikerforældre har lavere frafaldssandsynlighed på universitetet end individer, som ikke har akademikerforældre
2. Frafallssandsynligheden er størst i starten af studietiden

Den første hypotese understøttes af RRA teorien og Tintos frafaldsmodel. De studerende hvis forældre ikke har en universitetsuddannelse, har de ikke samme incitament til at fuldføre bacheloruddannelsen, som studerende med akademikerforældre, der endnu ikke har elimineret risikoen for nedadgående mobilitet. Videre har studerende med akademikerforældre i højere grad tilegnet sig uddannelsesrelevante evner gennem deres opvækst. Dette kan være med til at sikre en mere vellykket social og faglig integration på universitetsuddannelsen og derved en større vedholdenhed.

Den anden hypotese kan forklares ud fra Tintos frafaldsmodel og den dynamiske fortolkning af humankapitalteorien. Jf. Tinto sker frafald som et resultat af *mødet* mellem individet og uddannelsesinstitutionen, og frafaldssandsynligheden antages derfor at være størst i starten af studietiden. Jf. den dynamiske økonomiske beslutningsmodel vurderer individerne løbende, hvorvidt de skal fortsætte eller afbryde uddannelsen, da de inden studiestart antages at have imperfekt information. De lærer gradvist om egne evner og om uddannelsens afkast og omkostninger. Den studerende lærer for første gang om egne evner i relation til uddannelsen ved den første karaktergivning, som sker enten halvvejs inde eller i slutningen af 1. semester¹. Frafaldssandsynligheden antages derfor at være størst i starten af studietiden, hvor den studerendes læringen om uddannelsen og egne evner er størst.

¹Afhængigt af om uddannelsen følger blok- eller semesterstruktur.

4 Litteratur gennemgang

I dette afsnit præsenteres den eksisterende internationale og danske frafaldslitteratur. Frafaldslitteraturen er relativt omfattende, og der fokuseres derfor udelukkende på studier, som omhandler frafald på universitetet. Først fremlægges den eksisterende litteratur med fokus på forklaringsfaktorer for frafaldssandsynligheden på universitetet. Dernæst fremlægges den eksisterende litteratur med fokus på forældrebaggrunds betydning for frafald. Litteratur gennemgangens formål er at sætte problemformuleringen i perspektiv til den eksisterende litteratur og klargøre, hvordan denne afhandlings analyse bidrager til den eksisterende frafaldslitteratur.

4.1 Årsager til frafald

I dette afsnit fokuseres der på, hvilke faktorer der i litteraturen er fremlagt som forklaring på frafald. Der betragtes både individuelle baggrundskarakteristika og faktorer, der relaterer sig til studiemiljøet og det at studere. I frafaldslitteraturen er det vigtigt at skelne mellem betydningen af baggrundskarakteristika i dansk kontekst og øvrige landes. Det bero på, at betydningen af økonomiske faktorer er minimeret gennem gratis uddannelse i det danske uddannelsessystem samt eksistensen af SU-ordning [Smith et al., 2020, Andersen, 2018].

I det danske frafaldsstudie af Smith et al. [Smith et al., 2020] undersøges hvilke baggrundskarakteristika, der bedst forklarer frafald. De benytter en multipel cox-regression med clustered standardfejl på uddannelsesudbud. Deres datagrundlag består af studerende på Aarhus Universitets humanistiske fakultet. Studiet viser, at særligt typen af den adgangsgivende ungdomsuddannelse, adgangsgivende karaktergennemsnit og alder ved optag har stor betydning for frafald. Kønsforskelle i frafald forsvinder, når der kontrolleres for andre baggrundskarakteristika. De kontrollerer ikke for forældrebaggrund i cox-regressionen. Det understøttes af flere danske frafaldsstudier, at særligt det adgangsgivende karaktergennemsnit er af stor betydning for frafaldssandsynligheden [Smith et al., 2020, Uddannelses- og Forskningsministeriet, 2008, Ulriksen et al., 2015].

Qvortrup et al. [Qvortrup et al., 2018] undersøger betydningen af studiemiljøfaktorer for frafald gennem et litteraturstudie af international og dansk empirisk forskning. Mange af de faktorer som de finder har størst betydning for frafaldet, antages at variere på tværs af universiteter og uddannelser. Deres litteraturstudie viser, at de faktorer af størst betydning er det sociale og faglige aspekt af studiestarten, studenterdemografi, læringsfællesskaber samt relation og støtte fra undervisere og studievejledningstilbud.

I en analyse af EVA [Danmarks Evalueringsinstitut, 2017] undersøges det nærmere, hvilke aspekter af studiestarten, som spiller en rolle for frafaldet under det første halve år. Ved brug af survey- og registerdata viser de ved en fixed-effects linear probability model, at studerende med negative studiestartsoplevelser har større frafaldssandsynlighed det første halve år. Ved at undersøge forskellige aspekter af studiestarten finder de, at jo længere studiestartsforløbene er, jo lavere frafaldssandsynlighed har de studerende. Helt centralt for frafaldssandsynligheden er det sociale aspekt af studiestarten. Selv når der kontrolleres for andre konkrete aspekter af studiestarten har det sociale aspekt en selvstændig og signifikant betydning for det tidlige frafald. I analysen undersøges flere aspekter af studiestarten, som ikke har signifikant betydning

for det tidlige frafald. Det gælder den faglige og praktisk rammesætning², betydningen af introduktion til studieteknik, dannelse af studiegrupper og alkoholforbrug i studiesammenhænge.

Flere udenlandske studier beskæftiger sig med, hvordan frafald påvirkes af læring af egne evner gennem karakterpræstation [Cunha et al., 2005, Carneiro et al., 2003, Arcidiacono et al., 2016, Stinebrickner and Stinebrickner, 2012]. To af disse studier er præsenteret i det teoretiske afsnit 3.1. Det amerikanske studie af Stinebrickner & Stinebrickner [Stinebrickner and Stinebrickner, 2012] undersøger, hvorvidt frafald i løbet af det første studieår kan tilskrives, at de studerende får indsigt i deres egne akademiske evner gennem karaktergivning. Ved modelsimulering med data på amerikanske førsteårsstuderende på det amerikanske Berea College finder de, at omkring 40 pct. af førsteårsfrafaldet kan henføres til, at de studerende får indsigt i deres akademiske evner. Ligeledes undersøger Arcidiacono et al. [Arcidiacono et al., 2016] betydningen for frafald ved at individets lærer om egne evner gennem karakterer. De får samme resultat som Stinebrickner & Stinebrickner: en betydelig del af frafaldsbeslutningen kan tilskrives læring af egne evner. Det er vigtigt at være opmærksom på, at begge studier er baseret på amerikanske studerende, som modsat danske studerende oplever høj egenbetaling på universitetsuddannelser.

4.2 Betydningen af forældrebaggrund for frafald

Efter at have præsenteret hvilke faktorer - udover forældrebaggrund - der har betydning for frafald ifølge den eksisterende litteratur, fokuseres der nu på betydningen af forældrebaggrund.

I det britiske studie af Vignoles & Powdthavee [Vignoles and Powdthavee, 2009] undersøges det, hvorvidt universitetsstuderende fra ressourcetsvage hjem har højere frafaldssandsynlighed. Studiet er motiveret af at undersøge om frafald i højere grad tilskrives svage akademiske forudsætninger eller forældrebaggrundsrelaterede faktorer. Grundet manglende data på forældres uddannelse- og indkomstniveau benytter de, hvorvidt den studerende var berettiget til skolens gratis frokostordning som proxy for lavindkomst familier. Ud fra en probitmodel med clustered standardfejl på universitetsniveau konkluderer de, at der eksisterer et signifikant socioøkonomisk skel i frafaldet på britiske universiteter. En del af dette sociale skel kan forklares af forskelle i akademiske præstationer. Resultatet understøttes af Smith & Naylor [Smith and Naylor, 2001], som ved binomial regression ligeledes konkluderer, at frafaldssandsynligheden er signifikant og markant højere for britiske universitetsstuderende fra lavere socioøkonomisk baggrund. De to studiers resultater understøttes af det danske litteraturstudie af Larsen et al. [Larsen et al., 2013], der ved 16 forskellige europæiske studier fastslår, at der er solid evidens for en signifikant effekt af forældres uddannelsesniveau på frafaldssandsynligheden på universitetsniveau.

Majoriteten af forskningen af frafaldsårsager er amerikansk, men disse studier er svært relevante i dansk kontekst grundet den store brugerbetaling på universitetsuddannelser i USA. De to fremlagte britiske studier tager begge udgangspunkt i data fra før den større britiske uddannelsesreform i 2012, der øgede brugerbetalingen på videregående uddannelse markant. Resultaterne er derfor mere relevante i dansk kontekst end resultater baserede på uddannelsessystemer med en større grad af brugerbetaling.

I dansk kontekst understøtter Ulriksen et. al [Ulriksen et al., 2015], at forældrebaggrund har en signifi-

²Hvorvidt den studerende har oplevet at få en god introduktion til uddannelsens faglige indhold og krav og forventninger er blevet kommunikeret tydeligt.

kant betydning for frafaldssandsynligheden på universitet. Forældrebaggrund måles ved forældrenes uddannelsesniveau og analysen er afgrænset til studerende på STEM-uddannelser. Ved cox-regressionsanalyse viser de, at forældres uddannelsesniveau sammen med tidligere akademiske præstationer er de stærkest forklarende variable for frafaldssandsynligheden for studerende på STEM-universitetsuddannelser. Dette understøttes yderligere af en analyse udarbejdet af Idégruppen for at bryde den sociale arv på universitetsuddannelserne [Videnskabsministeriet, 2008]. Ved brug af logistisk regression finder de, at frafaldssandsynligheden på første studieår er større, jo lavere uddannede forældrene er. Dog finder de også, at andre forhold som uddannelsesvalg og adgangsgivende karaktergennemsnit, bidrager væsentligt mere til at forklare forskelle i frafaldssandsynligheden end forældrebaggrund gør.

I det danske frafaldsstudie af Smith et al. [Smith et al., 2020] undersøger de, foruden deres cox-regressionen præsenteret i afsnit 4.1, betydningen af forældres uddannelsesbaggrund for mødet med det faglige niveau og studiemiljøet. På baggrund af survey-data og en ANOVA-analyse, finder de, at forældres uddannelse ikke påvirker opfattelsen af det faglige niveau blandt de frafaldne studerende. Dog oplever studerende med lavtuddannede forældre i højere grad at få mindre støtte og feedback fra deres undervisere. Smith et al. mener, at en plausibel forklaring herpå kan være, at studerende med lavtuddannede forældre ikke genkender den støtte og feedback de modtager, da de ikke har lært at afkode omgangsformen i universitetsverden. Deres forklaring stemmer overens med teorien præsenteret i afsnit 3.3, som gennemgår, hvorfor studerende fra ikke-uddannelsesvante hjem kan føle sig mere fremmede og have svære ved at tyde de akademiske koder.

4.3 Bidrag til den eksisterende litteratur

Foruden at undersøge eksistensen af en social skævhed i de studerendes frafald på universitetet, bidrager denne afhandlings analyse ved at undersøge, om betydningen af forældrebaggrund for frafaldet *ændres* i løbet af studietiden. Det kan have stor betydning ved en eventuel politisk intervention, om der eksisterer en tidsmæssig afhængighed i betydningen af forældrebaggrund for frafald. Videre bidrager analysen til den eksisterende danske frafaldslitteratur med fokus på social skævhed ved, at analysen ikke er afgrænset til enkelte universiteter eller hovedområder, som det er tilfældet for [Ulriksen et al., 2015] og [Smith et al., 2020]. Afhandlingen er videre med til at understrege, hvor centralt et politisk fokus på, at studerende møder universitetsverden med forskellige baggrunde, er for den sociale retfærdighed³.

³Social retfærdighed skal i denne sammenhæng forstås som, at det kun er forhold, som individets selv har indflydelse på, der bør påvirke hvordan vedkommende klarer sig i samfundet [Roemer, 2002]

5 Data

I dette afsnit præsenteres analysens datagrundlag samt de variable, som er valgt til at indgå i analysens økonomiske model. Analysens datagrundlag består af registerdata på individniveau fra Danmarks Statistik og KOT under Uddannelses- og Forskningsministeriet. Dataadgangen til Danmarks Statistiks registerdata er opnået gennem Danmarks Evalueringsinstitut, EVA.

5.1 Populationsafgrænsning

Den samlede undersøgelsespopulation omfatter studerende, som blev optaget på en universitetsbacheloruddannelse gennem KOT i 2016 og 2017 med sommerstart. Det endelige datagrundlag består af 42.675 individer for hvem, den nødvendige data er tilgængelig.

I både 2016 og 2017 blev mere end 60.000 studerende optaget på en videregående uddannelse. Optaget for 2016 og 2017 er sammenlagt for at undgå årsspecifikke udsving i frafald eller faktorer korreleret hertil. Studenterpopulationen er afgrænset til sommerstartere for at sikre en mere homogen studenterpopulation. Datagrundlaget er videre afgrænset til at fokusere på de studerende med optag på en bacheloruddannelse på et af Danmarks otte universiteter.

Analysens fokus er universitetssektoren, da den sociale skævhed i uddannelsessystemet er mest udtalt på universitetsniveau [Nielsen et al., 2015]. Effekten af sociale baggrundsforhold på frafaldssandsynligheden formodes derfor at være størst i universitetssektoren, og derfor af størst interesse i et politisk perspektiv. Det er nødvendigt at afgrænse datagrundlaget til én uddannelsessektor grundet de store strukturelle forskelle mellem uddannelsessektorerne mht. studenterpopulation og uddannelsesspecifikke faktorer relateret til frafald. Alternativt kunne analysen være udført separat for de tre største sektorer inden for videregående uddannelse, men dette er fravalgt grundet afhandlingens omfang.

5.2 Anvendte empiriske variable

5.2.1 Den afhængige variabel

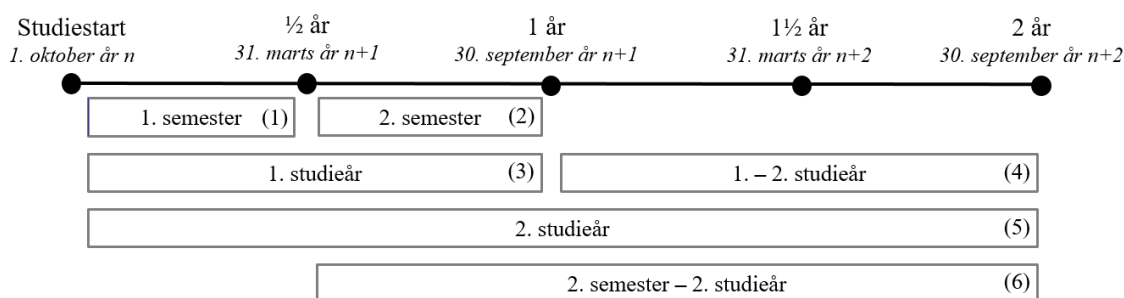
Modellens afhængige variabel er binær og beskriver, hvorvidt den studerende er faldet fra [= 1] eller stadig indskrevet på uddannelsen [= 0]. I denne analyse forstås frafald ved, at en påbegyndt uddannelse afbrydes inden uddannelsen er formelt fuldført. Der bliver ikke sondret mellem om de studerende påbegynder en ny uddannelse efter frafald eller ej. I afsnit 8.3 undersøges det nærmere, hvad de studerende foretager sig efter frafald.

Den specifikke dato, som fremgår af Danmarks Statistiks registerdata for, hvornår uddannelsen er afbrudt, skal ikke tillægges for stor analytisk betydning. Der er typisk et tidsrum mellem, at den studerende reelt frafalder uddannelsen og at den studerende aktivt afmelder sig uddannelsen eller at det administrativt registreres. Det kan bl.a. skyldes et økonomisk incitament, da SU fortsat udbetales, indtil den studerende afmelder sig uddannelsen. Uddannelsesinstitutioner registrerer typisk frafald i forbindelse med semesterskifte, hvor studerende ikke møder op til eksaminer, ikke tilmelder sig nye kurser eller registreres som optaget på en ny uddannelse. Denne antagelse understøttes empirisk af figur 9 i appendiks

A.2, som viser, hvordan den registrerede frafaldsdato for de frafaldne studerende i datagrundlaget fordeler sig. Histogrammet argumenterer for, at frafald hyppigst registreres i forbindelse med semesterskift, da der ses en høj densitet i registreret frafald omkring september måned.

Tidspunktet for frafald operationaliseres derfor ikke i kontinuert tid, men i diskret tid med inddelinger ved semesterskift. I analysen er fokus frafald inden for de to første studieår, som inddeles i seks forskellige perioder afgrænset ved semesterskift. For at kunne undersøge om der er empirisk belæg for, at forældrebaggrunds betydning for frafaldssandsynligheden ændres i løbet af studietiden, opereres der med seks forskellige variationer af den afhængige variabel. De seks forskellige afhængige variable beskriver frafaldssandsynligheden i seks forskellige perioder af de første to studieår. Den eneste forskel mellem de seks variable er altså, hvilken periode de måler om den studerende er frafaldet eller ej.

Figur 1: Perioder frafaldssandsynligheden estimeres



Note: Kursiv skrift angiver nedslagsdatoen, som deler tidsperioderne. n beskriver året for den studerendes studiestart, $n = 2016, 2017$.

Figur 1 viser de perioder frafaldssandsynligheden estimeres for. Den første periode beskriver frafaldet i løbet af det første halve år, 1. semester, og løber fra studiestart til 31. marts året efter. Studiestart defineres som 1. oktober, hvilket er en dato, der ofte benyttes i frafaldsanalyser for at frasortere studerende, som er blevet optaget på uddannelsen, men reelt ikke møder op⁴. Næste periode dækker det næste halve år, 2. semester, og løber fra 1. april til 30. september året efter studiestart. Perioden som beskriver det samlede førsteårsfrafald løber derved fra studiestart til 30. september året efter, hvor det samlede andetårsfrafald løber til 30. september to år efter studiestart. Perioden som dækker frafaldet mellem 1. og 2. studieår dvs. 3. til 4. semester løber fra 1. oktober året efter studiestart til 30. september to år efter studiestart. Den sidste periode beskriver andetårsfrafaldet med 1. semester udeladt dvs. fra 1. april året efter studiestart til 30. september to år efter studiestart. For periodeinddelingen er der medtaget en ekstra måned efter semesterskift under antagelsen om en forsinket administrativ registrering ift., hvornår det egentlig frafald indtræffer. Videre var ønsket at inddele året i to lige store perioder.

Variationen i frafaldsandelene for de seks perioder frafaldssandsynligheden estimeres for undersøges nærmere. Baseret på det samlede datagrundlag viser tabel 1 antal studerende, som er frafaldet i perioderne angivet i figur 1, og hvilken andel det udgør af det samlede datagrundlag. I løbet af de første to studieår er 27,5 pct. af de 42.675 studerende i datagrundlaget faldet fra på uddannelsen. Jf. en opgørelse fra Uddannelses- og Forskningsministeriet var andetårsfrafaldet mellem 25 og 29 pct. for universitetsbachelor-

⁴Kun 6 individer undlades i datagrundlaget ved at benytte 1. oktober i stedet for 1. september som studiestart. Havde studiestart været 1. september var datagrundlaget altså 42.681 individer i stedet for 42.675.

studerende med optag 2012-2015 med et gennemsnit på 27,25 pct. [Uddannelses- og Forskningsministeriet, 2018]. 19,3 pct. af de studerende i datagrundlaget er frafaldet i løbet af det første studieår og de resterende 8,2 pct. er faldet fra mellem 1. og 2. studieår. Det er et velkendt resultat, at størstedelen af studiefrafald sker i løbet af det første studieår [Uddannelses- og Forskningsministeriet, 2018]. Fokuseres der på frafaldet i løbet af henholdsvis 1. og 2. semester, ses det, at størstedelen af førsteårsfrafaldet sker i løbet af 2. semester. Det bemærkes, at frafaldet i løbet af 1. semester kun er marginalt højere end det samlede frafald mellem 1. og 2. studieår. Tabel 1 understøtter, at der er tilstrækkelig variation i frafald på tværs af de seks periodeinddelinger. For at sikre, at regressionsresultatet ikke er drevet af, at nogle perioder har større variation i frafald end andre perioder, er det andet studieår ikke inddelt i semestre, som det ses for det første studieår.

Tabel 1: Frafall for de seks perioder, optag 2016 og 2017

	Periode for frafall	Antal	Andel, pct.
(1)	1. semester	3.537	8,3
(2)	2. semester	4.705	11,0
(3)	1. studieår	8.242	19,3
(4)	1. - 2. studieår	3.490	8,2
(5)	2. studieår	11.732	27,5
(6)	2. semester - 2. studieår	8.195	19,2

Note: Andel af det samlede datagrundlag, som er frafaldet i den vilkårlige periode. Tallene i parentes i første kolonne referer til perioderne i figur 1.

Der gøres opmærksom på, at det ikke er muligt at skelne mellem selvvalgt frafall og frafall grundet tvangsudskrivning fra universitet⁵. Ud fra den tilgængelige data er frafallsvariablen begrænset af, at årsagen til uddannelsen afbrydes, ikke er identificeret. Ideelt havde frafallsvariablen udelukkende beskrevet sandsynligheden for selvvalgt frafall, da dette taler bedre ind i afhandlingens problemstilling som centrerer social skævhed i *uddannelsesvalg*.

5.2.2 Kontrolvariable

De anvendte kontrolvariable i den økonometriske analyse inddeles i at beskrive de studerendes individuelle karakteristika, uddannelsesmæssige baggrund og forældrebaggrund.

Individuelle karakteristika

Der vælges at kontrollere for den studerendes køn, alder og etniske herkomst. Kønsforskelle i uddannelsesvalg er et fælles resultat for mange danske studier, mens resultaterne er varierende i frafallslitteraturen [Troelsen, 2011]. Nyere dansk forskning finder dog, at der ikke er signifikant kønsforskel i frafallssandsynligheden på universitetet [Smith et al., 2020]. Køn inddrages i modellen som en binær variabel, hvor værdien 1 angiver kvinde og 0 angiver mand. Videre kontrolleres der for den studerendes alder ved tidspunkt for optaget. Alder indgår også i kvadreret form, da det antages, at alder har en aftagende effekt på frafallssandsynligheden. Den nyeste danske forskning finder, at alder har en ikke-lineær negativ sammenhæng med vedholdenhed

⁵Eksempelvis ved dumpede eksaminer eller manglende overholdelse af studieaktivitetskrav.

på universitetet [Smith et al., 2020]. Der kontrolleres for etnisk herkomst ved en binær variabel som tager værdien 1 for ikke-vestlige indvandrere og efterkommere og 0 ellers. Ikke-vestlig herkomst er defineret som indvandrere og efterkommere med oprindelsesland, som ikke er et EØS-land, Storbritannien, Andorra, Monaco, San Marino, Schweiz, Canada, USA, Australien eller New Zealand. Bestemmelsen af herkomst følger Danmarks Statistiks kategorisering af befolkningen i dansk oprindelse, indvandrere og efterkommere.

Uddannelsesmæssige baggrund

Den studerendes adgangsgivende karaktergennemsnit er valgt til at kontrollere for den studerendes faglige evner. Karaktergennemsnittet er et vægtet gennemsnit af eksamensresultater og årskarakterer over en 2-3 årige periode⁶ og eksklusiv eventuel karakterbonus. Det vurderes at være det mest retvisende tilgængelige mål for den studerendes akademiske evner. Resultatet fra en specifik eksamen ville i højere grad give et øjebliksbillede af den studerendes faglige præstation. For studerende med høje karaktergennemsnit kan denne variabel også opfange effekten af at være optaget på uddannelser med høje adgangskvotienter og derved en højere koncentration af studerende med højtuddannede forældre. Flere danske undersøgelser finder, at studerende med højtuddannede forældre får højere karaktergennemsnit på deres gymnasiale uddannelse end studerende med ufaglærte forældre [Thomsen, 2017, Danmarks Evalueringsinstitut, 2015]. I modelspecifikationen kontrolleres der for forskelle i studenterpopulationen på tværs ad uddannelsesudbud, se afsnit 7.2.2.

Det er et velkendt resultat, at jo højere adgangsgivende karaktergennemsnit den studerende har, jo lavere er frafaldssandsynligheden [Smith et al., 2020, Uddannelses- og Forskningsministeriet, 2008, Ulriksen et al., 2015]. Videre finder nyere danske undersøgelser, at typen af adgangsgivende ungdomsuddannelse har signifikant betydning for frafaldssandsynlighed [Danmarks Evalueringsinstitut, 2018, Smith et al., 2020]. Der kontrolleres derfor også for den studerendes adgangsgivende ungdomsuddannelse, hvor der skelnes mellem STX, HTX, HHX, HF og anden adgangsgivende uddannelse⁷, med STX som referencekategori.

Der kontrolleres videre for forhold relateret til studieoptaget, hvilket er indsamlet fra KOT-registeret. Den prioritet uddannelsen er ansøgt med, indgår ved en binær variabel, hvor værdien 1 angiver at være optaget på sin førstprioritet og 0 ellers. Der kontrolleres for optagelseskvote ved en kategorisk variabel, som angiver om den studerende er optaget på kvote 1 eller kvote 2, hvor referencekategorien er optag på en uddannelse med alle optaget.

Forældrebaggrund

Den primære variabel til at beskrive de studerendes forældrebaggrund er forældres højeste fuldførte uddannelsesniveau (HFU), da barnet var 15 år. Forælderen med det højeste uddannelsesniveau betegner barnets forældrebaggrund. Der skelnes mellem uddannelsesniveauerne grundskole, gymnasial uddannelse, erhvervsfaglig uddannelse, kort videregående uddannelse (KVU), mellemlang videregående uddannelse (MVU) og lang videregående uddannelse (LVU).

⁶Afhængigt af den adgangsgivende ungdomsuddannelse.

⁷Kan eksempelvis være en EUX eller en international studentereksamen.

Som alternativt mål for forældrebaggrund benyttes forældres realindkomstniveau, da barnet var 15 år. Det benyttede indkomstmål er summen af erhvervsindkomst, overførselsindkomst, formueindkomster og anden ikke klassificerbar indkomst, der kan henføres direkte til den enkelte person, og før fradrag af arbejdsmarkedsbidrag og særlig pensionsbidrag. Indkomsten er inflateret med forbrugerprisudviklingen med 2015 som basisår, da der er en tidsforskydning i, hvornår de studerende i datagrundlaget var 15 år. Indkomst inddrages i modellen i logaritmisk form for at opnå en mere intuitiv estimatfortolkning.

6 Deskriptiv analyse af data

I dette afsnit udføres der en deskriptiv analyse på baggrund af datagrundlaget. Tabel 2 præsenterer beskrivende statistik for det samlede datagrundlag, mens tabel 3 alene baserer sig på optaget sommer 2016, for at kunne beskrive fuldførelse og studieaktivitet tre år efter studiestart. Figur 2 er igen baseret på det fulde datagrundlag.

Tabel 2: Beskrivende statistik for datagrundlaget $N=42.675$

Variable	Middelværdi	Standardafvigelse
Individuelle karakteristika		
Alder	20,956	3,34
Kvinde	0,514	0,50
Ikke-vestlig herkomst	0,074	0,26
Uddannelsesmæssige baggrund		
Adgangsgivende karaktergennemsnit	8,159	2,05
STX	0,678	0,47
HTX	0,089	0,28
HHX	0,130	0,34
HF	0,100	0,30
Anden adgangsgivende uddannelse	0,003	0,05
Optaget på førsteprioritet	0,825	0,38
Optaget på kvote 1	0,660	0,47
Optaget på kvote 2	0,115	0,32
Uddannelse med alle optaget	0,224	0,42
Forældrebaggrund		
Grundskole	0,036	0,19
Gymnasial uddannelse	0,026	0,16
Erhvervsfaglig uddannelse	0,284	0,45
Kort videregående uddannelse	0,063	0,24
Mellemlang videregående uddannelse	0,328	0,47
Lang videregående uddannelse	0,263	0,44
Mors realindkomst (i tusinde kr.)	409,569	320044,20
Fars realindkomst (i tusinde kr.)	600,741	974953,50

Den beskrivende statistik viser, at en stor andel af datagrundlaget for analysen består af STX-studerende med dansk eller vestlig herkomst i start-20'erne med et over middel karaktergennemsnit fra den adgangsgivende eksamen og der ses en meget svag overrepræsentation af kvinder. Langt størstedelen er optaget på kvote 1 og deres førsteprioritet. Der ses en overrepræsentation af studerende, hvis forældre har en erhvervsfaglig uddannelse, MVU eller LVU som HFU. Fordelingen i forældres uddannelsesniveau blandt de studerende undersøges nærmere i næste tabel.

Tabel 3 viser fordelingen i forældrebaggrund blandt studerende i datagrundlaget med optag i 2016. Fordelingen vises for studieoptaget og videre for studerende, som har fuldført eller er indskrevet på studiet tre år senere, hvilket er den normerede studietid for en bacheloruddannelse. Tabellen beskriver kun studerende med optag i 2016, da data for fuldførelse og aktivitet tre år efter studiestart endnu ikke er tilgængeligt for studerende optaget i 2017. Som det ligeledes fremgik for det samlede datagrundlag i tabel 2 har størstedelen af de studerende optaget i 2016 forældre med en erhvervsfaglig uddannelse, MVU

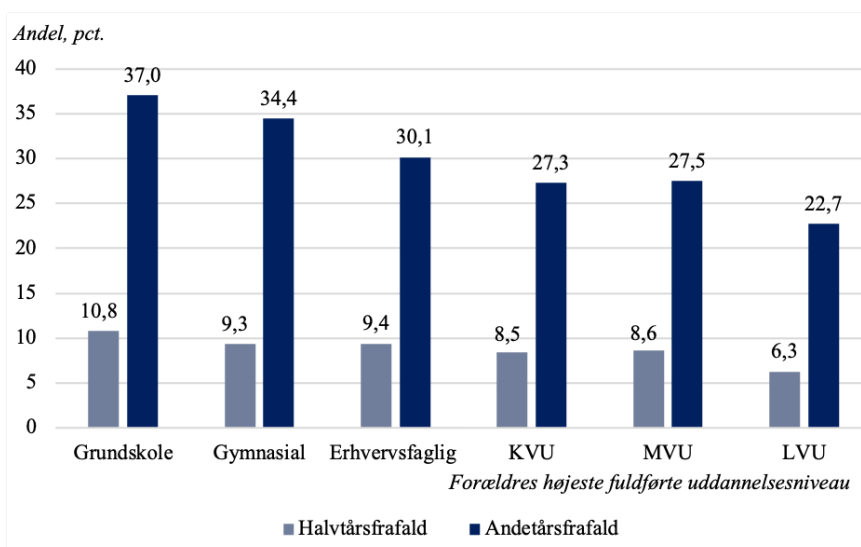
Tabel 3: Optag og fuldførelse fordelt på optagne studerendes forældres højeste fuldførte uddannelsesniveau

Andel, pct	Forældres højeste fuldførte uddannelsesniveau						Total
	Grundskole	Gym.	Erhvervsfaglig	KVU	MVU	LVU	
Optag	3,9	2,6	28,5	6,3	32,8	25,9	100,0
Fuldført eller aktiv	3,2	2,5	26,8	6,4	33,6	27,7	100,0

Note: Fuldførelse og aktivitet er målt 30. september 2019 dvs. tre år efter studiestart, hvor aktive studerende antages at fuldføre bacheloruddannelsen, som er normeret til tre år. Baseret på de studerende i datagrundlaget med optag i 2016.

eller LVU. Kun 3,9 pct. har forældre med grundskole som HFU. Sammenlignes der med de studerende, som fuldfører bacheloruddannelsen, ses der en øget overrepræsentation af forældre med en MVU og LVU, mens andelen af forældre med grundskole og erhvervsfaglige uddannelser falder. Tabel 3 indikerer, at den sociale skævhed i studerendepopulationen på universiteternes bacheloruddannelser forstærkes fra optag til fuldførelse gennem en social skævhed i frafaldet.

Figur 2: Frafald fordelt på optagne studerendes forældres højeste fuldførte uddannelsesniveau



Note: Halvtårsfrafald er andel frafaldne studerende målt 31. marts året efter studiestart og andetårsfrafald er andel frafaldne studerende målt 30. september to år efter studiestart.

Figur 2 viser en stærk systematik i frafaldsandelene på universitetet og forældrenes uddannelsesniveau. Jo kortere uddannelse de studerendes forældre har, jo flere falder fra. Tallene baserer sig på det fulde datagrundlag. Frafaldsandelen et halvt år efter studiestart er 10,8 pct. for de studerende, hvis forældre har en grundskole som HFU. Tilsvarende falder kun 6,3 pct. fra blandt de studerende med mindst én forældre med en LVU. Den procentvise afvigelse i halvtårsfrafaldet mellem de to studentergrupper er 71 pct. To år efter studiestart er 37,0 pct. af studerende, hvis forældre har grundskole som HFU faldet fra, hvor tilsvarende gælder for 22,7 pct. af de studerende med mindst én akademikerforældre. Det svarer til en procentvis afvigelse på 63 pct. mellem de to forældrebaggrunde. Fra tabel 3 og figur 2 ses en social skævhed i universitetsoptaget, og en tendens til, at flere studerende med højtuddannede forældre fuldfører den uddannelse de er optaget på, mens flere studerende med lavere uddannede forældre falder fra. Videre er denne sociale kløft i frafaldet svagt større ved halvtårsfrafaldet sammenlignet med andetårsfrafaldet.

7 Økonometrisk metode

I dette afsnit præsenteres den økonometriske teori bag logitmodellen samt begrundelsen for valget af en logitmodel til at estimere frafaldssandsynligheden. Slutteligt præsenteres den valgte modelspecifikation samt modellens udfordringer med inference.

7.1 Logitmodellen

Formålet med denne analyse er at estimere, hvordan forældrebaggrund påvirker sandsynligheden for frafald på universiteternes bacheloruddannelser, og om denne påvirkning ændres gennem studietiden. Da der opereres med frafald som en binær variabel, vælges der en binær responsmodel. Modellen estimeres for seks forskellige afhængige variable, som beskriver sandsynligheden for frafald i forskellige perioder af de første to studieår.

En linear probability model (LPM) er en ofte anvendt binære responsmodel. Modellen har dog nogle iboende ulemper såsom at kunne forudsige sandsynligheder udenfor det tilladte interval $[0, 1]$, samt modellens linearitet betyder, at de forklarende variables marginale effekter er konstante. Udfordringerne ved LPM kan afhjælpes ved at anvende mere sofistikerede ikke-lineære binære responsmodeller såsom probit- og logitmodellerne. Valget mellem disse to ikke-lineære responsmodeller er ikke afgørende for de empiriske resultater, da estimererne kun varierer marginalt [Wooldridge, 2016]. For begge modeller gælder, at responsandsynligheden er givet ved:

$$Pr(y = 1 | \mathbf{x}) = G(\mathbf{x}\boldsymbol{\beta}), \quad (1)$$

hvor $\mathbf{x}\boldsymbol{\beta}$ betegner det fulde sæt af forklarende variable og parametre. For logitmodellen er $G(\cdot)$ en logistisk funktion og den kumulative fordelingsfunktion følger en standard logistisk fordeling, hvilket er givet ved:

$$G(\mathbf{x}\boldsymbol{\beta}) = \frac{\exp(\mathbf{x}\boldsymbol{\beta})}{1 + \exp(\mathbf{x}\boldsymbol{\beta})} \quad (2)$$

Der gælder, at $G : \mathbb{R} \rightarrow [0;1]$. Ikke-lineariteten i $G(\cdot)$ sikrer, at de estimerede respons-sandsynligheder er indenfor det tilladte interval for sandsynligheder, hvilket gør logitmodellen væsentligt mere attraktiv end LPM, hvor fittede sandsynligheder kan være < 0 og > 1 . Probit- og logitmodellerne varierer mht. antagelsen om, hvilken kumulativ fordelingsfunktion $G(\cdot)$ følger og derved standardfejlenes fordeling. I denne afhandlings empiriske analyse benyttes der en logitmodel.

Logitmodellen kan afledes fra en underliggende latent variabel model, hvor y^* er en uobserveret variabel:

$$y^* = \mathbf{x}\boldsymbol{\beta} + e \quad \text{og} \quad y = \mathbf{1}[y^* > 0] \quad (3)$$

Her er $\mathbf{1}[\cdot]$ en indikatorfunktion, som tager værdien 1 ved $y^* > 0$ og værdien 0 ved $y^* \leq 0$. Det antages, at e følger en kontinuert standard logistisk fordeling og er uafhængig af \mathbf{x} . Det antages videre, at e er symmetrisk fordelt omkring 0, hvilket medfører, at $G(e) = 1 - G(-e)$. Givet disse antagelser bag (3) kan

respons-sandsynligheden y udledes:

$$\begin{aligned}
 Pr(y = 1 | \mathbf{x}) &= Pr(y^* > 0 | \mathbf{x}) \\
 &= Pr(e > -\mathbf{x}\boldsymbol{\beta} | \mathbf{x}) \\
 &= 1 - G(-\mathbf{x}\boldsymbol{\beta}) \\
 &= G(\mathbf{x}\boldsymbol{\beta}),
 \end{aligned}$$

hvilket vi genkender som udtryk (1) og husker, at $G(\cdot)$ er givet ved udtryk (2) hvilket sikrer, at de iboende ulemper ved LPM ikke længere er til bekymring [Wooldridge, 2016, Nielsen, 2015].

Maximum Likelihood Estimation

En logitmodel estimeres ved Maximum Likelihood Estimation (MLE). MLE er, under nogle generelle antagelser, den mest asymptotiske efficiente estimator.

Antaget, at $G(\cdot)$ følger en standard logistisk funktion, da er tæthedsfunktionen for y_i betinget på x_i for et datagrundlag bestående af n observationer, givet ved:

$$\begin{aligned}
 f(y | \mathbf{x}_i; \boldsymbol{\beta}) &= [G(\mathbf{x}_i; \boldsymbol{\beta})]^{y_i} [1 - G(\mathbf{x}_i; \boldsymbol{\beta})]^{1-y_i} \\
 &= \left(\frac{\exp(\mathbf{x}_i\boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i\boldsymbol{\beta})} \right)^{y_i} \left(1 - \frac{\exp(\mathbf{x}_i\boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i\boldsymbol{\beta})} \right)^{1-y_i} \\
 &= \left(\frac{\exp(\mathbf{x}_i\boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i\boldsymbol{\beta})} \right)^{y_i} \left(\frac{1}{1 + \exp(\mathbf{x}_i\boldsymbol{\beta})} \right)^{1-y_i}
 \end{aligned} \tag{4}$$

Sumproduktet af udtryk (4) er kendt som sample likelihood funktionen, $L_n(\boldsymbol{\beta})$. Log-likelihood funktionen opnås ved at tage log af sample likelihood funktion:

$$\begin{aligned}
 \mathcal{L}(\boldsymbol{\beta}) &= \log \left[\prod_{i=1}^n \left(\frac{\exp(\mathbf{x}_i\boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i\boldsymbol{\beta})} \right)^{y_i} \left(\frac{1}{1 + \exp(\mathbf{x}_i\boldsymbol{\beta})} \right)^{1-y_i} \right] \\
 &= \sum_{i=1}^n [y_i(\mathbf{x}_i\boldsymbol{\beta}) - \log(1 + \exp(\mathbf{x}_i\boldsymbol{\beta}))]
 \end{aligned} \tag{5}$$

Logit Maximum Likelihood estimatet, $\hat{\boldsymbol{\beta}}$, maksimerer $\mathcal{L}(\boldsymbol{\beta})$ således, at det observerede data gøres så sandsynligt som muligt givet den antaget model [Wooldridge, 2016, Nielsen, 2015].

MLE har nogle ganske attraktive asymptotiske egenskaber dvs. egenskaber som gælder når n konvergerer mod uendelig. Når $n \rightarrow \infty$ giver Store tals lov, at MLE er konsistent: $\hat{\boldsymbol{\beta}}_n \xrightarrow{P} \boldsymbol{\beta}_0$, hvor $\boldsymbol{\beta}_0$ er den sande værdi. Videre giver den centrale grænseværdi-sætning, at MLE er asymptotisk normalfordelt når $n \rightarrow \infty$: $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \xrightarrow{d} N(0, \Omega_\beta)$. Slutteligt er MLE asymptotisk efficient: alle andre konsistente og asymptotisk normalfordelte estimators har en asymptotisk varians større eller lig med MLEs asymptotiske varians [Nielsen, 2015].

Disse nævnte egenskaber gælder under nogle generelle antagelser. Det antages, at de forklarende variable er eksogene, responsvariablen er binær og følger derved en Bernoulli-fordeling samt modellens

funktionelle form er korrekt specificeret. Derudover antages det, at observationerne er identisk fordelt og uafhængige af hinanden. Eksogeniteten af x_i diskuteres i afsnit 7.2.1. Antagelsen om responsvariablen fordeling kan nemt argumenteres for er overholdt. I denne analyse beskriver $y_i = 1$ hændelsen, hvor den studerende afbryder sin uddannelse, og $P(y_i = 1|x)$ er sandsynligheden for frafald givet værdien af de forklarende variable, der inkluderes i modellen. Da y_i kun kan antage værdien 1 og 0 gælder, at responsvariablen følger en Bernoulli-fordeling. Videre antages det, at modellens funktionelle form følger udtryk (2). Dette er svært at validere empirisk, men det medfører ikke betydningsfulde forskelle i de empiriske resultater, om den funktionelle form i stedet var specificeret som en probitmodels [Wooldridge, 2016].

Hvorvidt den sidste antagelse om uafhængige observationer er overholdt er mere diskutabel. For at antagelsen er opfyldt skal der gælde, at en studerendes sandsynlighed for frafald ikke afhænger af de andre studerendes sandsynlighed for frafald. Det virker dog plausibelt, at en studerendes beslutning om frafald kan være påvirket af medstuderendes frafald. Denne påvirkning kan eksempelvis ske gennem læsegrupper, hvor én studerendes beslutning om frafald påvirker de andre i læsegruppen. Dette udfordrer antagelsen om uafhængige observationer, men da denne afhængighed ikke er systematisk, vil problemet være af mindre karakter. En anden bekymring for antagelsen er, at det samme individ kan optræde to gange i datagrundlaget, hvis vedkommen blev optaget i 2016, droppede ud og blev optaget igen i 2017. Denne afhængighed mellem observationerne tages der højde for ved cluster-robuste standardfejl på individ-niveau, se afsnit 7.2.2.

Fortolkning af logitmodellens estimater

For at kunne fortolke på størrelsen af effekten fra \mathbf{x} , de forklarende variable, på sandsynligheden for $P(y = 1 | \mathbf{x})$ i en logitmodel, er det nødvendigt at udregne de marginale effekter. Signifikans og retningen af effekten på $P(y = 1 | \mathbf{x})$ kan bestemmes ved β , mens effektens størrelse bestemmes ved de marginale effekter.

De marginale effekter udtrykker, hvor meget $P(y = 1 | \mathbf{x})$ ændrer sig som følge af en ændring i de forklarende variable. I tilfælde af en *kontinuert* variabel, x_k vil den marginale effekt udregnes således:

$$\frac{\partial P(y = 1 | \mathbf{x})}{\partial x_k} = \frac{\partial G(\mathbf{x}\boldsymbol{\beta})}{\partial x_k} = g(\mathbf{x}\boldsymbol{\beta})\beta_k, \quad (6)$$

hvor $g(z)$ er logitmodellens tæthedsfunktion, som er den partielt afledte af $G(\cdot)$ givet ved udtryk (2):

$$g(z) \equiv \frac{\partial G(z)}{\partial z} = \frac{\exp(z)}{[1 + \exp(z)]^2}$$

Hvis man ønsker at udregne den marginale effekt af ændringen fra 0 til 1 for en *diskret* variabel, x_j , gælder følgende:

$$G(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_{j-1} x_{i,j-1} + \hat{\beta}_j) - G(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_{j-1} x_{i,j-1}) \quad (7)$$

Givet udtryk (6) og (7) fremgår det, at de marginale effekter afhænger af værdierne af alle \mathbf{x} . Det medfører,

at effekterne for de forklarende variable ikke er ens for alle individer. En mulighed er at udregne de såkaldte *marginale effekter for gennemsnittet*. De udregnes som marginale effekter for hver forklarende variables gennemsnit i datagrundlaget. Ved denne tilgang opnår man den marginale effekt for den 'gennemsnitlige' person i datagrundlaget. En ulempe er, at gennemsnittet for diskrete variable ikke repræsenterer noget individ og ikke har fortolkningsmæssig betydning. I denne afhandlings empiriske analyse benyttes der derfor i stedet *den gennemsnitlige marginale effekt*⁸ til at fortolke på størrelsen af logitmodellens estimater. For en kontinuert variabel, x_k er den gennemsnitlige marginale effekt givet ved:

$$n^{-1} \sum_{i=1}^n \left[g(\mathbf{x}\hat{\boldsymbol{\beta}}) \hat{\beta}_k \right], \quad (8)$$

hvor n er antal af observationer i datagrundlaget. For den diskrete forklarende variabel, x_j , udregnes den gennemsnitlige marginale effekt af en ændring fra 0 til 1 ved:

$$n^{-1} \sum_{i=1}^n \left[G(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_{j-1} x_{i,j-1} + \hat{\beta}_j) - G(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_{j-1} x_{i,j-1}) \right] \quad (9)$$

De gennemsnitlige marginale effekter skal fortolkes som partielle effekter, da alle resterende forklarende variable end x_j i tilfælde af udtryk (9), holdes konstante ved deres observerede værdier [Wooldridge, 2016].

7.2 Den økonometriske model

Nedenfor opstilles logitmodellen, som estimeres ved MLE og hvis resultater præsenteres i afsnit 8.1:

$$\begin{aligned} Pr(\text{frafald} = 1 \mid \mathbf{x}) = G & \left(\beta_0 + \beta_1 K\text{vinde}_i + \beta_2 \text{Alder}_i + \beta_3 \text{Alder}_i^2 + \beta_4 \text{Ikke_vestlig}_i + \right. \\ & \beta_5 \text{Adg_gennemsnit}_i + \text{Adg_udd}_i \beta_6 + \beta_7 \text{Første_prio}_i + \beta_8 \text{Kvote1}_i + \\ & \left. \beta_9 \text{Kvote2}_i + \text{Forældre_HFU}_i \beta_{10} + \text{Udd_udbud}_i \beta_{11} \right) \end{aligned} \quad (10)$$

Der gælder, at $G(\cdot)$ følger udtryk (2) og \mathbf{x} repræsenterer modellens kontrolvariable, der følger definitionerne jf. afsnit 5.2.2. For korthedens skyld indgår Adg_udd_i og Forældre_HFU_i som vektorer af de kategoriske variable (referencekategorien udeladt). Variablen Udd_udbud_i er en vektor af uddannelses specifikke dummies, der kontrollerer for uobserveret heterogenitet på tværs af de specifikke uddannelsesudbud. Estimerne for β_{11} er udeladt, når regressionsresultaterne præsenteres.

Model (10) estimeres for seks forskellige afhængige variable, som beskriver frafaldssandsynligheden for forskellige tidsperioder i løbet af de første to studieår, gense her figur 1 for et overblik over de seks perioder.

7.2.1 Endogenitetsudfordringer

Der eksisterer et endogenitetsproblem, hvis frafaldssandsynligheden er påvirket af uobserverbare faktorer, der systematisk varierer på tværs af individer. Forklarende variable som er korreleret med disse

⁸Referes i litteraturen som både Average partial effect og Average marginal effect.

uobserverbare faktorer opfylder ikke antagelsen om eksogene forklarende variable.

Den studerendes personlighed kan eksempelvis antages at påvirke både frafaldssandsynligheden og flere forklarende variable. Det kan være gennem fx. faglig ambition, selvdisciplin, tilpasningsevne, socialt engagement og gåpåmod som sandsynligvis også er korreleret med forældrebaggrund. Videre er det en plausibel antagelse, at de enkelte uddannelsesudbud varierer på parametre, som påvirker de studerendes frafaldssandsynlighed. Der eksisterer derved en heterogeniteten i frafaldssandsynligheden på tværs af uddannelsesudbud, da de varierer både på studerendepopulation og faktorer som påvirker de studerendes vedholdenhed såsom forhold omkring studiestart, arbejdsbyrden, administrative forhold, adgangskvotienten mm. Særligt heterogeniteten i studenterpopulationerne er udfordrende at kontrollere for gennem forklarende variable, da uddannelserne varierer på, hvilke typer studerende de tiltrækker. Eksempelvis kan nogle uddannelser i højere grad tiltrække studerende med høj grad af ambition, selvdisciplin og arbejdsomhed, hvilket også er karaktertræk, der påvirker vedholdenhed på studiet. Der forsøges gennem det adgangsgivende karaktergennemsnit at tage højde for personlige karakteristika, som påvirker den studerendes faglighed, hvilket dog ikke er tilstrækkeligt som kontrol for heterogeniteten i uddannelsesudbud. Uddannelsesudbuddenes systematiske forskelle udgør en endogenitetsudfordring for estimationen af de studerendes frafaldssandsynlighed med fokus på forskelle drevet af forældrebaggrund.

7.2.2 Clustering og fixed-effects

I modellen benyttes der cluster-robuste standardfejl på individ-niveau, i , for at tage højde for korrelation *inden for* hver i og derved udfordringen ved opfyldelse af antagelsen om uafhængige observationer. For valid statistisk inference benyttes der clustered standardfejl på individniveau, da det samme individ kan optræde flere gange i datagrundlaget. Datagrundlaget består af universiteternes sommeroptag i 2016 og 2017, hvilket medfører, at et individ optaget i 2016 og frafaldet også kan fremgå i sommeroptaget 2017. Ved robuste clustered standardfejl tages der højde for, at observationer inden for hver gruppe i ikke overholder antagelsen om uafhængige observationer, som det er tilfældet med individer som optræder flere gange i datagrundlaget.

I modellen benyttes der fixed-effects på uddannelsesudbud-niveau for at tage højde for uobserverbar heterogenitet i frafaldssandsynligheden *mellem* uddannelsesudbud. Ved denne tilgang kontrolleres der for systematiske forskelle i studenterpopulationerne på tværs af uddannelser samt forskelle i frafaldet på tværs af uddannelser. Det antages, at studerende optaget på samme eller beslægtede uddannelser har en grad af systematisk samvariation på uobserverbare faktorer, som har betydning for frafaldssandsynligheden og andre forklarende variable i model (10). Videre er der betydelige forskelle i frafaldsandelene på tværs af de enkelte uddannelser. Der inkluderes derfor fixed-effects på uddannelsesudbud-niveau under antagelsen, at studerende som deler uobserverede baggrundsvariable selvselekterer ind på samme uddannelser og frafaldssandsynligheden varierer mellem bacheloruddannelserne.

Fixed-effects på de specifikke uddannelser er mest hensigtsmæssigt for at kontrollere for den uobserveret heterogenitet. Derved sammenlignes der i model (10) studerende optaget på samme specifikke uddannelsesudbud, og der tages højde for, at bestemte uddannelser i højere grad tiltrækker bestemte typer af studerende.

8 Resultater

Den økonometriske analyses hovedresultater præsenteres i dette afsnit med henblik på at besvare afhandlingens problemformulering: *Hvordan ændres betydningen af forældrebaggrund for frafaldssandsynligheden i løbet af studietiden?*. Videre kommenteres der på de teoretiske hypoteser opstillet i afsnit 3.4. Dernæst præsenteres to delanalyser. Første delanalyse undersøger ved estimation af subsamples eventuelle forskelle i betydningen af forældrebaggrund for frafaldssandsynligheden på tværs af de fem hovedområder. Den sidste delanalyse har til formål at kortlægge eventuelle sociale skævheder i, hvad de studerende som frafalder bacheloruddannelsen foretager sig efter frafaldet.

I nærværende analyse opereres der med et signifikansniveau på $\alpha = 0,05$. Variablenes signifikans vurderes på baggrund af z-statistikken, der asymptotisk følger en standard normalfordeling givet antagelserne i afsnit 7.1. Der gælder, at 97,5 percentilen i den kumulative fordelingsfunktion for en standard normalfordeling er 1,96, hvilket er den kritiske værdi, som nulhypotesen i følgende afsnit afvises på baggrund af [Nielsen, 2015]. Regressionstabellerne er præsenteret i appendiks A.1. Grundet specifikationen af model (10) med cluster-robuste standardfejl rapporteres der i regressionstabellerne den kvadreret Wald test i stedet for en likelihood-ratio test af den samlede models signifikans. Det skal bemærkes, at der grundet fixed-effects på de specifikke uddannelsesudbud, bliver udeladt nogle uddannelser i hver regression grundet multikollinearitet. Det fulde datagrundlag er 42.675 observationer og der går maksimalt 357 observationer tabt.

8.1 Hovedresultater

I dette afsnit præsenteres hovedresultaterne fra maximum-likelihood estimationen af den opstillet logit-model, model (10). Frafallssandsynligheden estimeres for de seks forskellige periodeinddelinger af de to første studieår, gense her figur 1. Formålet med regressionsanalysen er at undersøge betydningen af forældrebaggrund for frafallssandsynligheden under de to første studieår på universitetsbacheloruddannelserne. Videre fokuseres der på, hvordan betydningen af forældrebaggrund *ændres* i løbet af studietiden. Dette undersøges med de teoretiske hypoteser om, at frafallssandsynligheden er aftagende over studietiden og unge fra akademikerhjem har lavere frafallssandsynlighed på universitetet end andre studerende jf. afsnit 3.4.

Først præsenteres de prædikterede gennemsnitlige marginale effekter for forældres uddannelsesniveau for frafallssandsynligheden i model (10), hvor frafallssandsynligheden estimeres for de seks forskellige perioder. Dernæst præsenteres seks kvadrerede Wald test af forældrebaggrund-variablen for de seks forskellige perioder frafallssandsynligheden estimeres. Til slut undersøges det nærmere, hvor stærk betydningen af at have akademikerforældre er for frafallssandsynligheden ift. at have forældre med lavere uddannelsesniveauer.

Som tidligere nævnt i afsnit 7.1, er det de *gennemsnitlige marginale effekter*, der rapporteres i denne analyse. Det betyder, at ved fortolkningen af de marginale effekter af forældrebaggrund holdes alle andre kontrolvariable end forældres uddannelsesniveau konstant ved deres observerede værdier. De marginale effekter afhænger derved af karakteristikaene for alle individer i estimationens datagrundlag. Da grundskole

som HFU er referencekategorien for variabelen, som beskriver forældres uddannelsesniveau, fortolkes de gennemsnitlige marginale effekter af forældres HFU som: *den gennemsnitlig ændring i frafaldssandsynligheden på tværs af alle observationer når forældres uddannelsesniveau ændres fra grundskole til et andet vilkårligt uddannelsesniveau*. I figur 3 er de prædikterede gennemsnitlige marginale effekter præsenteret for MLE estimationerne af model (10) for de seks forskellige perioder. Regressionstabellen for alle seks regressioner er præsenteret i tabel 10 i appendiks A.1.

Som alternativ til figur 3 ses figur 8 i appendiks. Figur 8 viser ligeledes de prædikterede gennemsnitlige marginale effekter. Dog er frafaldssandsynlighederne ikke sat relativt til frafaldssandsynligheden for studerende med forældre med grundskole som HFU, som det ses i figur 3. Figur 8 viser de gennemsnitlige marginale frafaldssandsynligheder, hvis alle individer i datagrundlaget havde forældre med det givne uddannelsesniveau - alle andre karakteristika holdt konstant. På den vis gives et overblik over niveauforskelle i de prædikterede frafaldssandsynligheder på tværs af de seks perioder frafaldssandsynligheden er estimeret for.

Figur 3a viser, at for 1. semester er frafaldssandsynligheden - i gennemsnit på tværs af alle individer i datagrundlaget - 2,9 procentpoint lavere for studerende med mindst én akademikerforældre ift. hvis begge forældre har grundskole som HFU. Dette svarer til, at hvis alle individer i datagrundlaget har forældre med grundskole - alle andre karakteristika holdt konstant - har de i gennemsnit en frafaldssandsynlighed på 9,7 pct. i løbet af 1. semester. Tilsvarende er frafaldssandsynligheden 6,8 pct., hvis alle studerende har mindst én forældre med en LVU, se figur 8a i appendiks. De studerende hvis forældre har grundskole som HFU har altså 30 pct.⁹ højere frafaldssandsynligheden på 1. semester end studerende med akademikerforældre.

Forskellen i frafaldssandsynligheden mellem studerende med forældre med grundskole som HFU og akademikerforældre er signifikant ned til et 1%-signifikansniveau. Forskellen mellem grundskole og de resterende uddannelsesniveauer er insignifikant. Ses der på frafaldssandsynligheden under 2. semester har forældres uddannelsesniveau ingen signifikant betydning¹⁰.

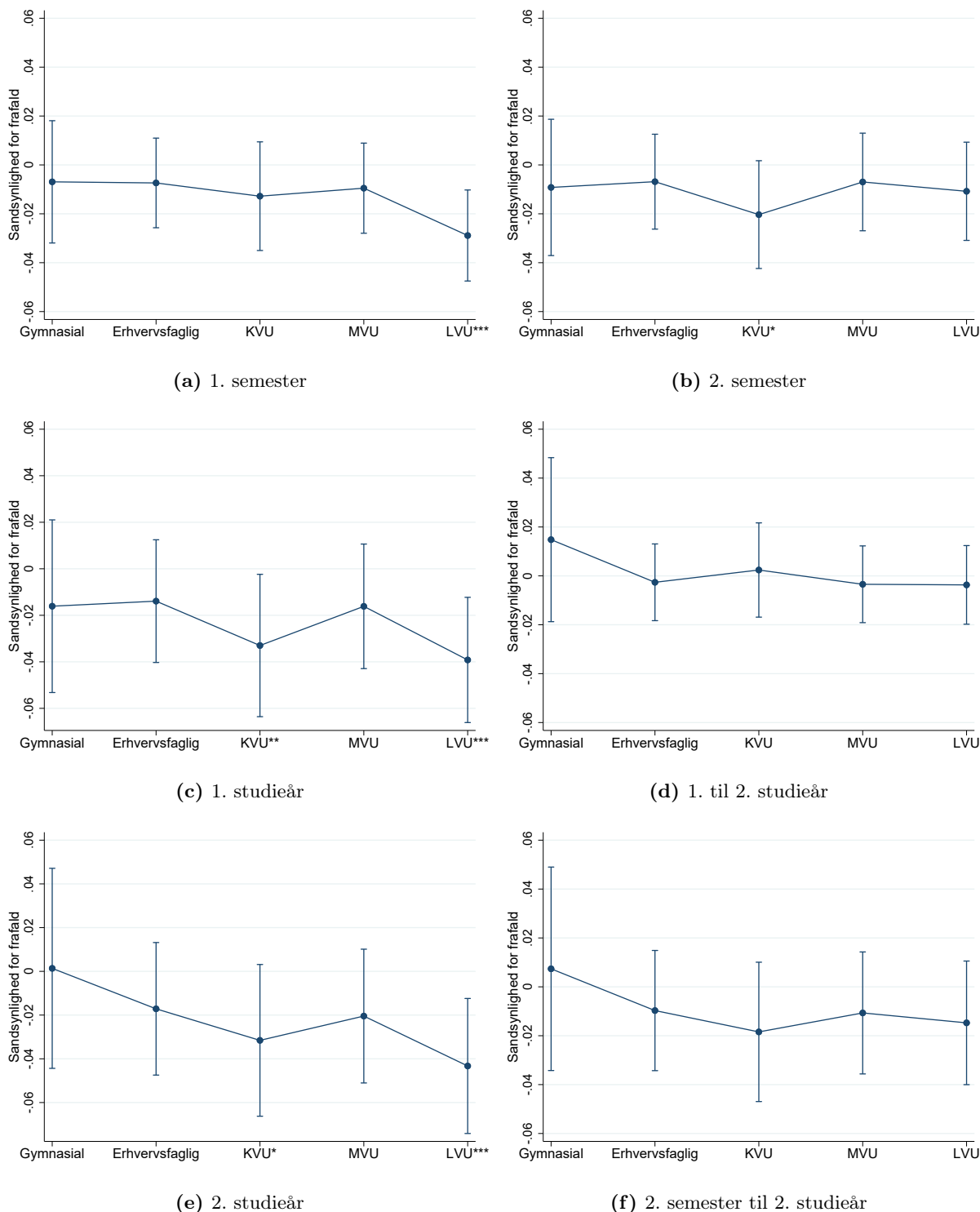
For det samlede 1. studieår er frafaldssandsynligheden signifikant lavere for studerende hvis forældre har en KVVU eller LVU relativt til, hvis de har grundskole som HFU. I gennemsnit og givet karakteristikaene for datagrundlaget, har de studerende med forældre med en KVVU 3,3 procentpoint lavere frafaldssandsynlighed end hvis begge forældre har grundskole som HFU. Hvis forældrene har en LVU er tilsvarende frafaldssandsynlighed 3,9 procentpoint lavere, se figur 3c. Hvis begge forældre har grundskole som HFU, da har den studerende, i gennemsnit, 21,5 pct. sandsynlighed for at falde fra i løbet af det samlede 1. studieår, mens tilsvarende er 17,6 pct., hvis den studerende kommer fra et akademikerhjem, se figur 8c i appendiks. For det efterfølgende studieår svarende til 3. og 4. semester - figur 3d - har det ingen signifikant betydning for frafaldssandsynligheden om forældrene har et højere uddannelsesniveau end grundskole eller ej. Ved at sammenligne de prædikterede frafaldssandsynligheder i figur 8c og 8d i appendiks, fremgår det tydeligt, at frafaldssandsynligheden for alle forældres uddannelsesniveauer er markant højere under det samlede 1. studieår ift. det efterfølgende studieår.

For de samlede to første studieår er frafaldssandsynligheden for studerende med forældre med grundskole

⁹ $\frac{9,7-6,8}{9,7} * 100 = 30,0$ pct.

¹⁰ Om forældre har grundskole eller en KVVU giver signifikant forskellige frafaldssandsynligheder på et 10%-signifikansniveau, hvilket klassificeres som insignifikant i denne analyse.

Figur 3: Prædikterede marginale sandsynligheder med 95% konfidensinterval - grundskole som referencekategori



Note: X-aksen viser forældres højeste fuldførte uddannelsesnivea, og y-aksen ændringen i frafaldssandsynligheden relativt til hvis forældrene havde grundskole som højeste fuldførte uddannelsesnivea. Der refereres til signifikansniveauet af forældres uddannelsesnivea relativt til hvis forældrene havde grundskole: *** $p < 0,01$, ** $p < 0,05$, * $p < 0,1$. Frafallssandsynlighederne er estimeret ved model (10) for de seks perioder præsenteret i figur 1.

og forældre med LVU forskellig på et 1%-signifikansniveau, se figur 3e. I gennemsnit på tværs af alle individer i datagrundlaget, er effekten af at have mindst én forældre med en LVU ift. hvis begge forældre har

grundskole en 4,3 procentpoint lavere frafaldssandsynligheden under de to første studieår. For tilsvarende periode men med 1. semester udeladt - dvs. 2.semester til 2. studieår - er frafaldssandsynligheden uafhængig af forældrenes uddannelsesniveau. Hvis 1. semester er indeholdt er der derimod en stærk signifikant forskel mellem studerendes frafaldssandsynlighed, hvis deres forældre har grundskole ift. en LVU. Dette resultat fremgår tydeligt ved at sammenligne figur 3e med 3f, hvor eneste forskel er, at 1. semester ikke er indeholdt i perioden frafaldssandsynligheden er estimeret for i figur 3f.

Foruden at være signifikant, så har forældres uddannelsesniveau også en relativ stor effektstørrelse i estimationen af frafaldssandsynligheden på 1. semester relativt til de andre kontrolvariable i modellen. Effektstørrelsen af at have forældre med LVU ift. at have forældre med grundskole er en 2,9 procentpoint lavere frafaldssandsynlighed. Den eneste anden signifikante kontrolvariabel, som har større effektstørrelse end forældre uddannelsesniveau er den studerendes adgangsgivende uddannelse. Hvis den studerende har HF som adgangsgivende uddannelse ift. STX har vedkommende 4,8 procentpoint højere frafaldssandsynlighed. Fokuseres der på det samlede første- og andetårsfrafald har forældres uddannelsesniveau stadig en af de største effektstørrelser, men nu har den studerendes optagelseskvote og -prioritet foruden adgangsgivende uddannelse større effektstørrelser end forældres uddannelsesniveau.

Resultaterne præsenteret i figur 3 indikerer, at forældres uddannelsesniveau kun har signifikant betydning for frafaldssandsynligheden på 1. semester. For at understøtte dette resultat, undersøges det nærmere om variabelen, som beskriver forældres uddannelsesniveau i model (10), er af statistisk signifikant betydning for frafaldssandsynligheden i de seks forskellige perioder¹¹.

Test af forældrebaggrund som signifikant prediktor af frafaldssandsynlighed

I de prædikterede sandsynligheder præsenteret i figur 3 er den statistiske signifikans af forældres uddannelsesniveau *relativ* til, hvis forældrene har grundskole som HFU. I dette afsnit undersøges forældrebaggrundvariablens overordnede signifikans i model (10) for de seks frafaldsperioder.

Det testes om variabelen, som beskriver forældrenes HFU, er en statistisk signifikant prediktor for de studerendes frafaldssandsynlighed givet de andre variable, der kontrolleres for i model (10). Dette testes ved en kvadreret Wald test. Grundet model (10) specifikation med cluster-robuste standardfejl benyttes den kvadreret Wald test i stedet for en likelihood-ratio test. Ved en likelihood-ratio test estimeres model (10) med og uden forældrebaggrund-variabelen, hvor faldet i likelihood-størrelsen er udtryk for, hvor meget mindre sandsynligt de observerede data er blevet. På samme vis beskriver den kvadreret Wald test sandsynligheden af den estimerede model med forældrebaggrund sammenholdt med modellen uden forældrebaggrund-variabelen. En kvadreret Wald test har under nulhypotesen en asymptotisk χ^2 -fordeling. Testens z test-størrelse evalueres på baggrund af en kritisk værdi på 11,07, hvilket er baseret på χ^2 -fordelingen med 5 frihedsgrader for et signifikansniveau på $\alpha = 0,05$.

Resultaterne af de seks kvadrerede Wald test er præsenteret i tabel 4. I model (10) er variabelen, som beskriver forældres uddannelsesniveau en stærk signifikant prediktor af sandsynligheden for frafald

¹¹Det er blevet undersøgt, om der findes en kønsforskel i betydningen af forældrebaggrund for frafaldssandsynligheden. Model (10) er blevet estimeret med et interaktionsled mellem forældres uddannelsesniveau og køn. Der blev ikke fundet en signifikant kønsforskel i, hvordan forældrebaggrund påvirker frafaldssandsynligheden.

Table 4: Kvadreret Wald test af forældres højeste fuldførte uddannelsesniveau

		<i>Kvadreret Wald test, $\chi^2(5)$</i>
	Periode for frafald	p-værdi z test-størrelse
(1)	1. semester	0,000 30,43
(2)	2. semester	0,414 5,02
(3)	1. studieår	0,000 24,24
(4)	1. - 2. studieår	0,818 2,22
(5)	2. studieår	0,009 20,80
(6)	2. semester - 2. studieår	0,653 3,30

Note: Tallene i parentes i første kolonne refererer til perioderne i figur 1. De kvadrerede Wald test er baseret på model (10), hvor frafaldssandsynligheden er estimeret for de seks perioder.

under 1. semester, 1. studieår og 2. studieår givet de andre kontrolvariable i model (10). Estimeres frafaldssandsynligheden derimod for 2. semester, 1. til 2. studieår og 2. semester til 2. studieår, er forældres uddannelsesniveau ikke en signifikant prediktor. Resultaterne for de kvadrerede Wald test understøtter tendensen fundet fra regressionsresultaterne: forældres uddannelsesniveau er kun af betydning for frafaldssandsynligheden i perioder, hvor 1. semester er indeholdt. Særligt interessant er at sammenligne Wald testen for de samlede to første studieår og 2. semester til 2. studieår, hvor eneste forskel er, at 1. semester ikke er indeholdt i sidstnævnte periode. For frafaldssandsynligheden i perioden mellem studiestart og to år frem er forældrebaggrund-variablen signifikant på et 1%- signifikansniveau. For samme periode men med 1. semester udeladt, er forældrebaggrund nu en insignifikant prediktor for frafaldssandsynligheden med en p-værdi på 0,653.

Nærværende analyse viser, at forældres uddannelsesniveau har signifikant betydning for frafaldssandsynligheden på 1. semester på en universitetsbacheloruddannelse. Frafaldssandsynligheden efter 1. semester er ikke afhængig af forældres uddannelsesniveau. Det er altså i starten af universitetsuddannelserne, at forældres uddannelsesniveau spiller en rolle for de studerendes frafald.

Analyse af betydningen af at have akademikerforældre

Det undersøges i dette afsnit, om den stærke signifikante forskel i frafaldssandsynligheden på 1. semester for studerende med akademikerforældre er drevet af, at der sammenlignes med studerende, hvis forældre har det laveste mulige uddannelsesniveau. Det undersøges ved at ændre referencekategorien for den kategoriske variabel for forældres uddannelsesniveau. Grundskole er valgt som referencekategori, da det giver den mest intuitive fortolkning at sammenligne med det laveste uddannelsesniveau. Det betyder dog også, at der for studerende med akademikerforældre sammenlignes studerende med to meget forskellige forældrebaggrunde. Det undersøges, om en signifikant lavere frafaldssandsynlighed for studerende med akademikerforældre er drevet af at være sammenlignet med forældre med det laveste mulige uddannelsesniveau, *eller* om studerende er mere vedholdende på deres universitetsuddannelse end andre studerende, hvis forældrene også har læst på universitetet.

I tabel 5 ses de prædikterede gennemsnitlige marginale effekter af frafaldssandsynligheden på 1. semester for forældrenes forskellige uddannelsesniveauer. Referencekategorien skifter mellem de forskellige uddannelsesniveauer, hvor estimerne i den første kolonne er med grundskole som referencekategorien.

Tabel 5: Prædikterede marginale sandsynligheder med diverse uddannelsesniveauer som referencekategori for forældres højeste fuldførte uddannelsesniveau, 1. semester frafaldssandsynlighed

<i>Forældres udd.niveau.</i>	<i>Referencekategori</i>					
	Grundskole	Gym.	Erhvervsfag.	KVU	MVU	LVU
Grundskole	-					
Gymnasial	-0,0069 [0,587]	-				
Erhvervsfaglig	-0,0074 [0,431]	-0,0004 [0,964]	-			
KVU	-0,0128 [0,259]	-0,0059 [0,610]	-0,0054 [0,470]	-		
MVU	-0,0095 [0,312]	-0,0026 [0,791]	-0,0021 [0,599]	0,0033 [0,669]	-	
LVU	-0,0289*** [0,002]	-0,0220** [0,026]	-0,0215*** [0,000]	-0,0161** [0,036]	-0,0194*** [0,000]	-

Note: p-værdier i [·]. Frafallssandsynligheden er estimeret for model (10) for 1. semester. Der refereres til signifikansniveauet af forældres uddannelsesniveau relativt til forældre med givet uddannelsesniveau som højeste fuldførte uddannelsesniveau (fremgår i fed skrift): *** $p < 0,01$, ** $p < 0,05$, * $p < 0,1$.

For overblikkets skyld er de marginale effekter til højre for diagonalen i matrixen udeladt, da matrixen er skævsymmetrisk¹². Tabel 5 viser, at de studerende, hvis forældre har en LVU, har signifikant lavere frafallssandsynlighed på 1. semester uagtet hvilken anden forældregruppe de holdes op imod. Selv relativt til studerende hvis forældre har en MVU, har studerende med forældre med en LVU i gennemsnit 1,9 procentpoint lavere frafallssandsynlighed på 1. semester, hvor tilsvarende er 2,9 procentpoint når grundskole anvendes som referencekategori. Det er et overraskende resultat, at frafallssandsynligheden er stærkt signifikant forskellig mellem studerende med forældre med LVU og MVU, der antageligvis ikke er så markant forskellige forældrebaggrunde. Alligevel er det af signifikant betydning for frafallssandsynligheden om én af forældrene er akademiker eller har en MVU.

Denne økonomiske analyses hovedresultater stemmer overens med de teoretiske hypoteser præsenteret i afsnit 3.4: De studerende hvis forældre har en LVU, har signifikant lavere frafallssandsynlighed end alle andre studerende, og frafallssandsynligheden er for alle studerende størst tidligt i studieforløbet. Analysens resultater indikerer videre, at forældres uddannelsesniveau kun spiller en rolle for de studerendes frafall på 1. semester. Som mulig forklaring, kan der inddrages samme teorier, som ligger til grund for de to teoretiske hypoteser. En intuitiv fortolkning kan være, at de studerende på baggrund af første tilbagemelding om egne evner, opdaterer deres vurdering af det forventet afkast og omkostninger knyttet til uddannelsen. Her er det plausibelt, at der eksisterer et socialt skel i de studerendes faglige selvtillid og deres motivation for at fuldføre uddannelsen efter de lærer om egne evner i relation til uddannelsen. Den manglende motivation til at fuldføre uddannelsen for særlig studerende med ufaglærte forældre kan understøttes af den sociologiske teori om relativ risikoaversion. I Tintos model for frafall vil resultatet kunne forklares af samspillet mellem den enkelte studerende og uddannelsesinstitutionen, hvor studerende med ikke-akademiker forældre i mødet med universitetsverden i højere grad oplever en negativ eller utilstrækkelig tilknytning til uddannelsen, hvilket bl.a. kan være gennem følelsen af ikke at høre til

¹²Matrixens transponerede er lig dens negative, $A^T = -A$.

på universitetet. Det kan være drevet af, at forældre med lange uddannelser i højere grad kultiverer uddannelsesrelevante evner hos deres børn. Den dynamiske humankapitalteori, RRA teorien og Tintos frafaldsmodel er gennemgået i afsnit 3.

8.2 Delanalyse: forskelle på tværs af hovedområder

I denne delanalyse undersøges det, om der eksisterer forskelle i forældrebaggrunds betydning for frafaldssandsynligheden på tværs af universiteternes fem hovedområder. Formålet er at kunne identificere for hvilke hovedområder problemstillingen om social skævhed i frafald er mest presserende, hvilket vil være af betydning ved eventuel politisk intervention.

Resultaterne fra analysen i afsnit 8.1 indikerer, at betydningen af forældres uddannelsesniveau udelukkende er statistisk signifikant for bachelorstuderendes frafaldssandsynlighed i løbet af 1. semester. Der koncentrerer derfor i denne delanalyse på frafaldet i løbet af 1. semester, da formålet er at afgøre, hvor i universitetssektor en eventuel politisk indsats er mest hensigtsmæssig. Datagrundlaget inddeler i subsamples efter det hovedområde den studerende er optaget på. De fem hovedområder følger universitetssektorens hovedområder jf. afsnit 2.1. Først kortlægges der ved en deskriptiv analyse de sociale forskelle i optag og frafald på tværs af hovedområderne. Slutteligt estimeres 1. semester frafaldssandsynligheden for model (10) for de fem subsamples.

Tabel 6 viser fordelingen i forældres HFU for optaget på de fem hovedområder for datagrundlaget. Optaget på humaniora har både den lavest andel af studerende fra akademikerhjem og den største andel med forældre, der har grundskole som HFU. Sundhedsvidenskab er det hovedområde, hvor studerende hvis forældre er ufaglærte (grundskole eller gymnasial uddannelse) fylder den største andel af hovedområdets optag. De absolutte tal for fordelingen i forældres uddannelsesniveau for optaget kan ses i tabel 14 i appendiks A.2.

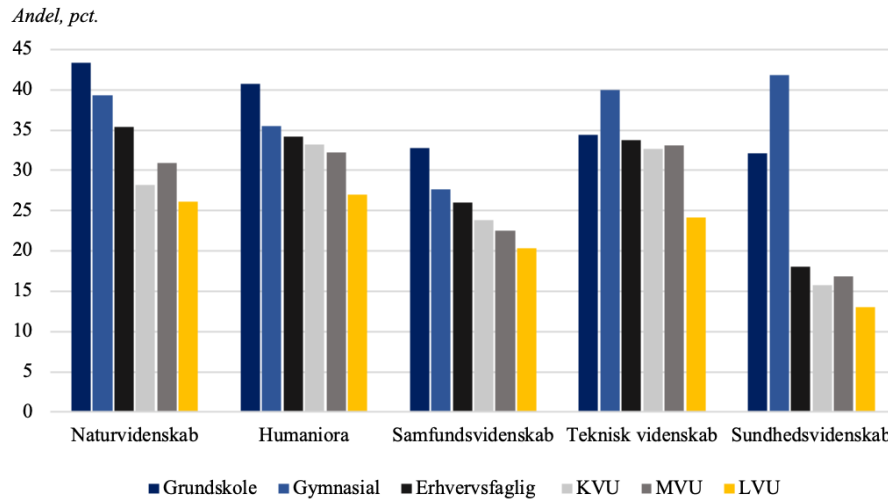
Tabel 6: Fordelingen i forældres højeste fuldførte uddannelsesniveau blandt optag fordelt på hovedområder

<i>Andel, pct.</i>	<i>Forældres højeste fuldførte uddannelsesniveau</i>						Total
	Grundskole	Gym.	Erhvervsfag.	KVU	MVU	LVU	
Naturvidenskab	3,3	2,5	27,5	5,8	32,8	28,0	100,0
Teknisk videnskab	2,9	1,9	24,7	6,6	33,7	30,2	100,0
Humaniora	4,1	2,6	30,3	6,3	35,1	21,6	100,0
Samfundsvidenskab	3,5	2,6	29,5	6,4	31,0	27,6	100,0
Sundhedsvidenskab	3,9	4,1	23,6	6,8	32,0	29,6	100,0

Figur 4 viser frafaldet opdelt på hovedområder og fordelt på forældres uddannelsesniveau. Frafaldsandelene er deskriptive dvs. der korrigeres ikke for andre baggrundsvariable. Af hensyn til risiko for identifikation af enkeltindivider kan frafaldet på 1. semester fordelt på forældres uddannelsesniveau ikke benyttes, da der er for få observationer for sundhedsvidenskab. Figur 4 viser derfor andetårsfrarafaldet. Det laveste samlede andetårsfrarafald er på sundhedsvidenskab, hvor 17,5 pct. af de studerende var frafaldet efter 2. studieår, mens det højeste andetårsfrarafald ses på humaniora, hvor tilsvarende andel er 32,2 pct. og dernæst naturvidenskab med 31,3 pct. frafald. For alle hovedområder med undtagelse af teknisk videnskab ses der en tendens til højere frafaldsandel, jo lavere uddannelsesniveau forældrene har. Forskellene i

frafaldsandelene for studerende med ufaglærte forældre - grundskole eller gymnasial uddannelse som HFU - relativt til højere uddannelsesniveauer er særdeles slående for det sundhedsvidenskabelige hovedområde. Sundhedsvidenskab er også det hovedområde, hvor selvsamme forældrebaggrunde udgjorde den største andel af det samlede optag, som primært består af studerende optaget på medicin-studiet.

Figur 4: Andetsårsfrafald fordelt på optagne studerendes forældres højeste uddannelsesniveau, opdelt på hovedområder



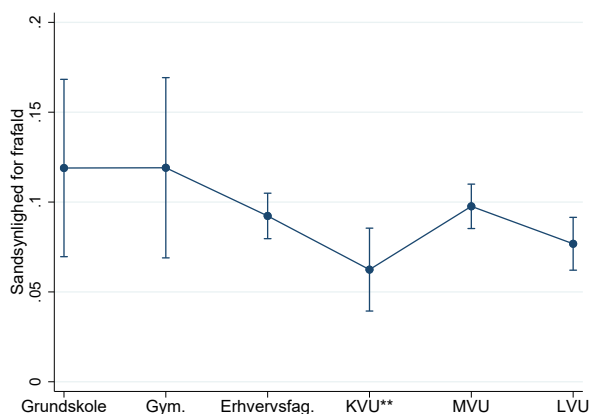
Note: Andetsårsfrafaldet er andel frafaldne studerende målt 30. september to år efter studiestart.

For at bedst muligt isolere betydningen af forældrebaggrund for frafaldssandsynligheden, estimeres model (10) for hver af de fem subsamples, hvor den afhængige variabel, $Pr(\text{frafald} = 1)$, beskriver frafaldssandsynligheden for 1. semester. Det er med undtagelse af det sundhedsvidenskabelige hovedområde, hvor frafaldssandsynligheden er estimeret for det samlede 1. studieår. Det skyldes, at sundhedsvidenskabs optagelse- og frafaldsmønstre udskiller sig fra de resterende fire hovedområder. Sundhedsvidenskab er det hovedområde med det mindste optag og de laveste frafaldsande. Det medfører en for lav variation i frafaldet under 1. semester til meningsfulde fortolkninger af estimationsresultater fra model (10). Figur 5 viser de prædikterede gennemsnitlige marginale frafaldssandsynligheder for hvert uddannelsesniveau for forældrene, mens regressionstabellerne for de fem hovedområder er præsenteret i tabel 11 i appendiks afsnit A.1.

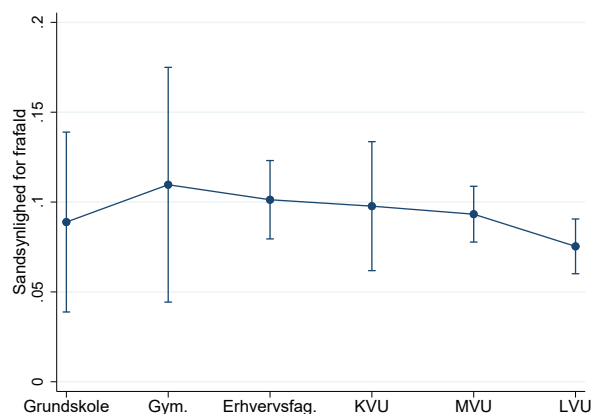
Frafaldssandsynlighederne i figur 5 skal fortolkes som frafaldssandsynligheden i gennemsnit, hvis alle i subsamplet havde forældre med det givet HFU - alle andre karakteristika holdt konstant ved deres observerede værdi. Fordelen ved, at de prædikterede gennemsnitlige marginale effekter for forældres HFU ikke sættes relativt til grundskole er, at niveauforskelle på tværs af hovedområder bliver mulige at aflæse. Her er det vigtigt at huske, at frafaldssandsynligheden for sundhedsvidenskab er for det samlede 1. studieår og ikke 1. semester som for de resterende hovedområder.

For naturvidenskab er det kun for forældre med KVVU, at de studerende i gennemsnit har en signifikant lavere frafaldssandsynlighed relativt til hvis forældrene, har grundskole som uddannelsesniveau. I gennemsnitlige - på tværs af alle studerende i subsamplet for naturvidenskab - har studerende med forældre med grundskole 11,9 pct. sandsynlighed for frafald i løbet af 1. semester. Har mindst én af forældrene en KVVU

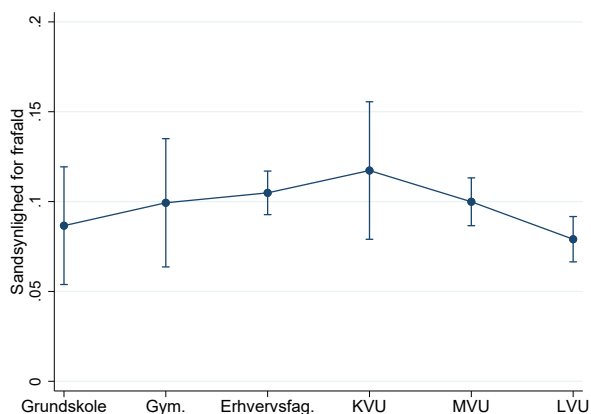
Figur 5: Prædikterede marginale sandsynligheder uden referencekategori med 95% konfidensinterval - opdelt på universiteternes hovedområder



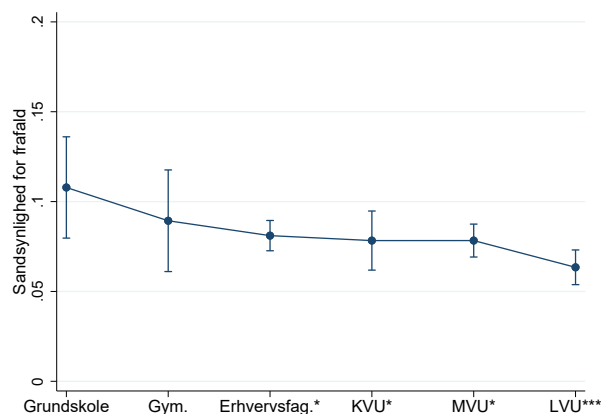
(a) Naturvidenskab



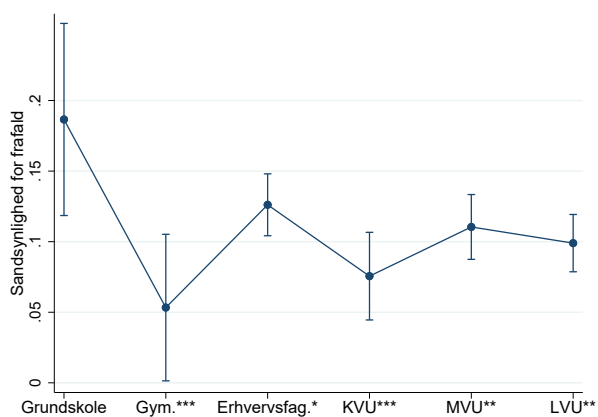
(b) Teknisk videnskab



(c) Humaniora



(d) Samfundsvidenskab



(e) Sundhedsvidenskab (1. studieår)

Note: X-akserne viser forældres højeste fuldførte uddannelsesniveau. Frafallssandsynligheden er estimeret for model (10) for 1. semester med undtagelse af sundhedsvidenskab, hvor frafallssandsynligheden er estimeret for det samlede 1. studieår. Der refereres til signifikansniveauet af forældres uddannelsesniveau relativt til hvis forældrene havde grundskole som højeste fuldførte uddannelsesniveau: *** $p < 0,01$, ** $p < 0,05$, * $p < 0,1$.

er frafallssandssynligheden 5,7 procentpoint lavere. For de bachelorstuderende på teknisk videnskab og humaniora ses der ingen signifikant forskel i frafallssandssynligheden om forældrene har grundskole eller et

højere uddannelsesniveau, se figur 5b og 5c.

Forældres uddannelsesniveau har en væsentligt større betydning for frafaldssandsynligheden på samfundsvidenskab. På tværs af alle bachelorstuderende i subsamplet for samfundsvidenskab, har studerende med mindst én akademikerforældre i gennemsnit 4,4 procentpoint lavere frafaldssandsynlighed end hvis begge forældre har grundskole som HFU. Forskellen er signifikant ned til et 1%-signifikansniveau. Forskellen i den gennemsnitlige frafaldssandsynlighed for studerende med forældre med en MVU relativt til grundskole har en p-værdi på 0,0504 og er derfor på grænsen til at være signifikant.

På sundhedsvidenskab ses der en kraftigt reduceret frafaldssandsynlighed for studerende hvis forældre har et højere uddannelsesniveau end grundskole, se figur 5e. For alle uddannelsesniveauer på nær erhvervsfaglig er frafaldssandsynligheden signifikant lavere end hvis forældrene har grundskole. De prædikterede frafaldssandsynligheder er for 1. studieår og ikke 1. semester og derfor større end ved de resterende fire estimationer.

Alternativt kunne model (10) i denne analyse være estimeret med fixed-effects på institutionsniveau i stedet for de specifikke uddannelsesudbud. Datagrundlaget for hver subsample er afgrænset til studerende på samme hovedområde, som derved studerer på beslægtede uddannelser. Da der er afgrænset til samme hovedområder, kan det antages, at den uobserverede heterogenitet mellem de studerendes frafaldssandsynlighed er større på tværs af universiteterne end de enkelte uddannelser. Det har dog ingen indflydelse på resultaterne om modellen estimeres med fixed-effects på uddannelsesudbud- eller institutionsniveau, og de er derfor kun præsenteret for fixed-effects på de specifikke uddannelsesudbud.

Tabel 7: Kvadreret Wald test af forældres højeste fuldførte uddannelsesniveau for hovedområderne

Hovedområde	Kvadreret Wald test, $\chi^2(5)$	
	p-værdi	z test-størrelse
Naturvidenskab	0,000	30,43
Teknisk videnskab	0,414	5,02
Humaniora	0,000	24,24
Samfundsvidenskab	0,818	2,22
Sundhedsvidenskab	0,009	20,80

Note: De kvadrerede Wald test er baseret på model (10), hvor frafaldssandsynligheden er estimeret for 1. semester med undtagelse af sundhedsvidenskab, hvor frafaldssandsynligheden er estimeret for 1. studieår.

Figur 5 viser, at forældrebaggrund har en signifikant betydning for frafaldssandsynligheden på det naturvidenskab og særligt på samfundsvidenskab og sundhedsvidenskab. Disse resultater understøttes af kvadrerede Wald test for hvert af de fem subsamples. For naturvidenskab, samfundsvidenskab og sundhedsvidenskab er forældres uddannelsesniveau en signifikant prediktor for frafaldssandsynligheden, mens forældrebaggrund ikke er af betydning for frafaldssandsynligheden på teknisk videnskab og humaniora. Resultaterne fra de kvadrerede Wald test er præsenteret i tabel 7. Denne delanalyse resultater indikerer, at det er mest hensigtsmæssigt med politisk fokus på det naturvidenskabelige hovedområder og særligt det samfundsvidenskabelige- og sundhedsvidenskabelige hovedområde. Dette resultat understøttes af forskningslitteraturen, som finder, at der er størst social ulighed i uddannelse på 'traditionelle fag' såsom medicin, statskundskab, jura og økonomi [Uddannelses- og Forskningsministeriet, 2008], hvilke er de uddannelser, som udgør størstedelen af optaget på henholdsvis det sundhedsvidenskabelige- og

samfundsvidenskabelige hovedområde.

Resultaterne i denne analyse skal tages med forbehold. Med henblik på, at sikre en tilstrækkelige variation i data er estimationen for sundhedsvidenskab for en anden periode end de resterende hovedområder. Størrelsen af det samlede optag og forældregrupperne varierer meget på tværs af de fem hovedområder, se tabel 14 i appendiks A.2. Det kan derfor diskuteres om resultaterne er drevet af forskelle i variationen på tværs af de fem subsamples. Det kan undersøges nærmere ved at medtage flere studieoptag i estimationen for at skabe et større datagrundlag og derved variation. Grundet afhandlingens omfang er dette fravalgt.

8.3 Delanalyse: de studerende som falder fra

Som det fremgik af den deskriptive analyse i afsnit 6 og blev bekræftet af den økonometriske analyse, ses der en social skævhed i frafaldet på universiteternes bacheloruddannelser. Problematikken ved et socialt skævt frafald forstærkes, hvis der også ses social slagside i hvilke studerende, der efter frafald forlader uddannelsessystemet og er uden for arbejdsstyrken. Dette kan sammenlignes med at indtræde i den såkaldte NEET-gruppe¹³, som betegner unge, som hverken er under uddannelse eller i beskæftigelse. Der gøres opmærksom på, at der definatorisk afviges fra Danmarks Statistiks NEET-gruppe, som er unge i alderen 16-24 år, som på opgørelsestidspunktet (den sidste uge i november) er uden for beskæftigelse og ikke under uddannelse samt har ikke været under uddannelse de foregående tre uger.

Den danske forskningslitteratur bekræfter en social slagside i risikoen for at ende i NEET-gruppen, hvor de unge ofte har en historik med frafald i uddannelsessystemet, hvilket oftest sker fra erhvervsuddannelser [Bolvig et al., 2019]. Det undersøges, om der findes en social skævhed i, hvad der sker efterfølgende med de studerende som frafalder. Fokus er på hvem, der påbegynder en ny uddannelse, hvilken uddannelsestype de påbegynder, og hvad dem som ikke påbegynder en ny uddannelse foretager sig.

I denne delanalyse tages der udgangspunkt i de 20.228 studerende i datagrundlaget med optag i 2016. Der fokuseres videre på de 4.086 studerende (svarende til 20,2 pct.) som var frafaldet under 1. studieår. Med en social vinkel undersøges det, om disse studerende er påbegyndt en ny uddannelse, som de ikke afbryder, samt om de studerende, der har forladt uddannelsessystemet, er lønmodtagere eller uden for arbejdsstyrken. Ultimo 2018¹⁴ opgøres deres tilknytningen til arbejdsmarkedet samt, hvorvidt de er påbegyndt en ny uddannelse, der ikke er blevet afbrudt. For at tillade tilstrækkelig tid til at observere om de unge evt. påbegynder ny uddannelse efter frafald, tages der udgangspunktet i 2016 optaget og førsteårsfrafaldet.

Tabel 8: Andel blandt frafald som påbegynder en ny videregående uddannelse der ikke afbrydes, fordelt på forældres højeste fuldførte uddannelsesniveau - ultimo 2018

<i>Andel, pct.</i>	<i>Forældres højeste fuldførte uddannelsesniveau</i>					
	Grundskole	Gym.	Erhvervsfag.	KVU	MVU	LVU
Påbegyndt ny uddannelse	67,8	66,7	69,7	72,2	62,8	64,3

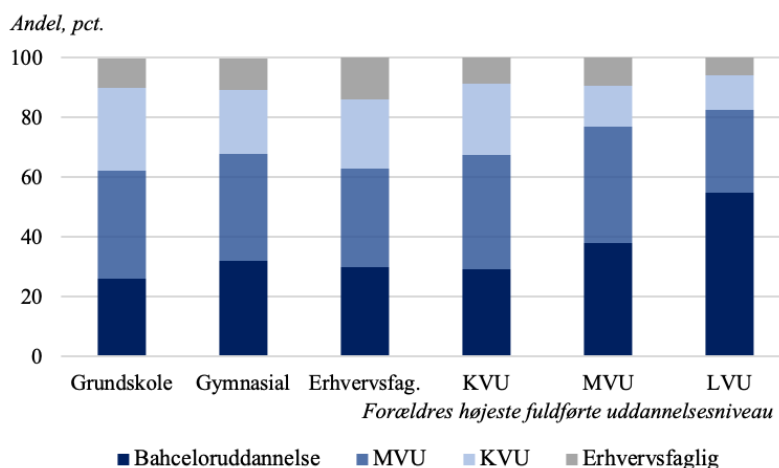
Note: Data dækker over de 4.086 studerende, der blev optaget i 2016 og frafaldt i løbet af 1. studieår. Andelen beskriver de studerende som efter frafald har påbegyndt en ny videregående uddannelse de enten har færdiggjort eller ikke har afbrudt. Opgjort ultimo 2018.

¹³Not in Education, Employment, or Training

¹⁴Tidspunkt for nyeste tilgængelige data fra Danmarks Statistiks arbejdsmarkedsregnskab.

Af de studerende som frafaldt i løbet af første studieår var 66,3 pct. påbegyndt en ny videregående uddannelse, som de enten havde færdiggjort eller var aktive på ultimo 2018. Der ses ikke en social skævhed i hvem, som påbegynder en ny uddannelse efter frafald. Den største andel som påbegynder en ny uddannelse efter frafaldet, er de studerende, hvis forældre har en KVVU som HFU, se her tabel 8.

Figur 6: Fordeling i uddannelsestype påbegyndt eller færdiggjort efter frafald, fordelt på forældres højeste fuldførte uddannelsesniveau - ultimo 2018



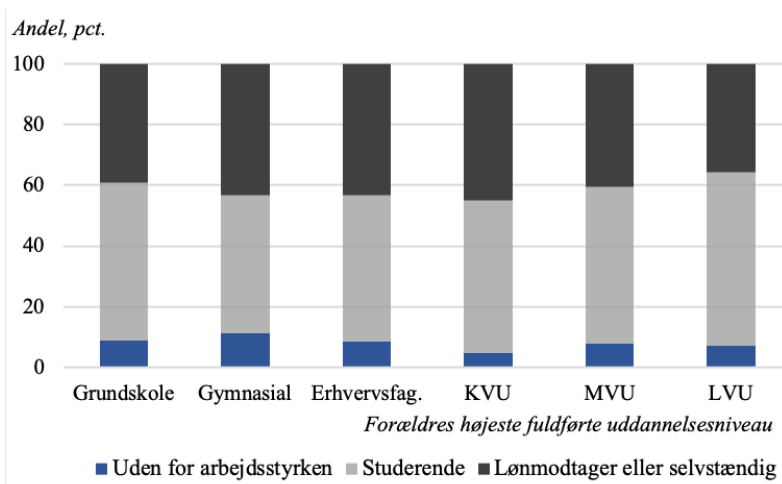
Note: Data dækker over de 2.706 studerende, der blev optaget i 2016 og frafaldt i løbet af 1. studieår og som har påbegyndt en ny videregående uddannelse de enten har færdiggjort eller ikke afbrudt. Opgjort ultimo 2018.

For de studerende som efter frafald påbegynder en ny uddannelse, der ikke afbrydes, viser figur 6 fordelingen i uddannelsestyper for den nye uddannelse. Figuren viser, at de studerende i højere grad vender tilbage til deres forældres uddannelsesniveau efter frafald på universitetet. Unge fra akademiske hjem påbegynder i langt højere grad en bacheloruddannelse på ny end unge med lavere uddannede forældre. Blandt de studerende med forældre med grundskole, som påbegynder en ny uddannelse, påbegynder 26,1 pct. af dem en bacheloruddannelse på ny. Dette er tilsvarende for 54,8 pct. af de studerende med mindst én akademikerforældre. Det ses også, at den største andel som påbegynder en erhvervsfaglig uddannelse efter frafaldet, er de studerende, hvis forældre selv har en erhvervsfaglig uddannelse. Tilsvarende er den største andel, 39,0 pct., som påbegynder en MVU efter frafald, de studerende med forældre med en MVU.

Figur 7 viser fordelingen i de frafaldne studerendes tilknytning til arbejdsmarkedet opgjort ultimo 2018. Andelen af aktive studerende i figur 7 er lavere end andelen præsenteret i tabel 8, da tabel 8 viser andelen, som har påbegyndt en ny uddannelse uden at afbryde den. Derved er studerende, som har færdiggjort den nye uddannelse også indeholdt i tabel 8. Der ses ikke en tendens til, at studerende med lavere uddannede forældre i højere grad er uden for arbejdsstyrken efter frafald. Generelt er det lave andele på tværs af alle forældrebaggrunde, som ender uden for arbejdsstyrken. Det kan bl.a. forklares af, at der tages udgangspunkt i unge som tidligere har været indskrevet på en universitetsuddannelse, og det er derved ikke de mest socialt udsatte unge, som det ellers oftest ses blandt NEET-gruppen.

Problematikken ved den sociale skævhed i frafaldet på universitetet forstrækkes ikke af nærværende analyse. Der ses ikke en tendens til, at det i højre grad er de unge med lavtuddannede forældre, som efter frafald forlader uddannelsessystemet eller ender uden for arbejdsstyrken. Analysen viser dog et socialt skel i hvilken uddannelsestype, der påbegyndes efter frafald. Der er en svag tendens til, at de studerende

Figur 7: Fordeling i de frafaldne studerendes tilknytning til arbejdsmarkedet, fordelt på forældres højeste fuldførte uddannelsesnivea - ultimo 2018



Note: Data dækker over de 4.086 studerende, der blev optaget i 2016 og frafaldt i løbet af 1. studieår. Opgjort ultimo 2018.

efter frafald på universitetet søger mod samme uddannelsesnivea som deres forældre. Denne tendens understøtter den sociologiske teori om relativ risikoaversion præsenteret i afsnit 3.2.

9 Robusthedsanalyse

I dette afsnit undersøges robustheden af afhandlingens hovedresultater. Robustheden analyseres ved ændringer i specifikationen af model (10).

9.1 Indkomst som mål for forældrebaggrund

I hovedanalysen benyttes forældres uddannelsesniveau som mål for de studerendes forældrebaggrund. Alternativt kan forældrenes indkomstniveau benyttes til at opfange betydningen af forældrebaggrund. I model (11) er den kategoriske variabel, som beskriver forældrenes uddannelsesniveau, erstattet med kontinuerte variable, som beskriver henholdsvis morens og farens log-transformerede realindkomstniveau. Dette er den eneste ændring fra model (10). Indkomst-variablen er nærmere beskrevet i afsnit 5.2.2.

$$\begin{aligned} Pr(\text{frafald} = 1 \mid \mathbf{x}) = G & (\delta_0 + \delta_1 \text{Kvinde}_i + \delta_2 \text{Alder}_i + \delta_3 \text{Alder}_i^2 + \delta_4 \text{Ikke_vestlig}_i + \\ & \delta_5 \text{Adg_gennemsnit}_i + \text{Adg_udd}_i \delta_6 + \delta_7 \text{Første_prio}_i + \delta_8 \text{Kvote1}_i + \\ & \delta_9 \text{Kvote2}_i + \delta_{10} \text{Mor_indkomst_log}_i + \delta_{11} \text{Far_indkomst_log}_i + \\ & \text{Udd_udbud}_i \delta_{12}) \end{aligned} \quad (11)$$

Det er et velkendt resultat, at højere uddannelsesniveau fører til højere indkomstniveau [Skaksen, 2018, Pihl, 2009]. Forældres uddannelses- og indkomstniveau er derfor tæt korreleret. Af den grund inkluderes begge variable for forældrebaggrund ikke i samme model, da de vil tage forklaringseffekt fra hinanden¹⁵. Model (11) estimeres ved MLE for de seks perioder, hvor morens- og farens indkomstniveau nu er inkluderet som kontrolvariable. Datagrundlaget indskrænkes fra 42.675 til 42.537 individer for hvem data på begge forældres indkomst i barnets 15. leveår er tilgængelig. Regressionsresultaterne er præsenteret i tabel 13 i appendiks A.1.

Maximum-likelihood estimationen af model (11) viser, at morens indkomstniveau ikke spiller nogen rolle for frafaldssandsynligheden, mens farens indkomstniveau har en meget lille, men signifikant betydning for frafaldssandsynligheden efter 1. studieår. Sandsynligheden for frafald i løbet af de to første studieår falder i gennemsnit med 0,1 procentpoint ved en 10% stigning i farens indkomstniveau i barnets 15. leveår - alle andre observationer holdt konstant. Farens indkomstniveau er modsat forældres uddannelsesniveau ikke kun signifikant i de perioder, hvor 1. semester er indeholdt. Farens indkomstniveau er derimod udelukkende signifikant i de perioder, som indeholder det 2. studieår, og er insignifikant i de perioder, der kun beskriver frafald i løbet af det 1. studieår.

Regressionsresultaterne indikerer, at frafald først bliver afhængigt af forældrenes (farens) indkomstniveau senere i studietiden. En plausibel forklaring kan være, at det først er senere i studieforløbet, at studerende bliver slået ud af økonomisk modgang. Hvis forældrene har et højt indkomstniveau, kan det mindske barnets frafaldssandsynlighed gennem eksempelvis en mere stabil boligsituation og en roligere studietid. Økonomisk støtte fra forældrene giver mulighed for at opsigte studiejob for at øge fokus på

¹⁵Inkluderes både forældres uddannelses- og indkomstniveau i samme model ændres signifikansen ikke af forældres uddannelsesniveau, men størrelsen af den marginale effekt bliver mindre. Farens indkomstniveau skifter fra at være signifikant til insignifikant, når forældres uddannelsesniveau også er inkluderet i modellen.

studiet hvis nødvendigt. Hvis den studerende gennem længere tid hyppigt skal skifte bolig og konstant har stor usikkerhed forbundet med deres boligsituation, kan det føre til højere frafaldssandsynlighed på lang sigt. En analyse fra EVA viser, at studerende, som bekymrer sig meget over deres boligsituation, er mere tilbøjelige til at frafalde end andre studerende [Danmarks Evalueringsinstitut, 2019]. Det antages her, at forældrenes indkomstniveau påvirker den studerendes boligsituation positivt. Dette er blot én plausibel forklaring på, hvorfor forældres indkomstniveau først har betydning for det sene frafald, men der kan ligge mange forskellige mekanismer bag.

Forældrebaggrunds betydning for frafald ændres, når forældres indkomstniveau anvendes som mål for forældrebaggrund som alternativ til forældres uddannelsesniveau. Afhandlingens hovedresultater indikerer, at forældres uddannelsesniveau kun har betydning for det tidlige frafald (på 1. semester). Anvendes forældres indkomstniveau, findes et andet resultat. Frafalet er uafhængigt af morens indkomst, mens farens indkomstniveau udelukkende har signifikant betydning for det sene frafald (efter 1. studieår). Det tidlige frafald mindskes ved at have akademikerforældre, antaget, at forældrenes kulturelle værdier og normer er med til at sikre en mere succesfuld social og faglig integration ved mødet med universitetsverden. På længere sigt har forældrenes uddannelsesniveau ingen betydning for frafalet, men forældres indkomstniveau spiller nu en rolle. Det antages at bl.a. være gennem muligheden for en mere rolig og stabil studietid uden at økonomisk modgang stjæler fokus fra studiet.

9.2 Fixed-effects på alternativt uddannelsesudbud-niveau

Der er i model (10) inkluderet fixed-effects på de specifikke uddannelsesudbud. Formålet er at kontrollere for uobserveret variation mellem uddannelser drevet af forskelle i uddannelses- og institutionsspecifikke karakteristika og selvselektion i, hvilken type studerende der blive optaget på uddannelserne. Fixed-effects på de specifikke uddannelsesudbud kontrollerer bedst for denne heterogenitet i frafaldssandsynligheden på tværs af uddannelsesudbud. Det har den ulempe, at der i estimationen udelades nogle observationer grundet multikollinearitet på enkelte uddannelser. Det er ikke af større bekymring, da der maksimalt mistes 357 observationer ud af 42.675. Som robusthedstjek erstattes fixed-effects på de specifikke uddannelsesudbud med fixed-effects på hovedområde-institution-niveau, hvor ingen observationer går tabt. Det betyder, at der ikke længere sammenlignes studerende optaget på samme specifikke uddannelsesudbud eksempelvis Økonomi, KU. Der sammenlignes nu studerende optaget på samme hovedområde på samme universitet eksempelvis Samfundsvidenskab, KU.

Tabel 9 viser de kvadrerede Wald test af forældrebaggrund-variablen i model (10), hvor fixed-effects er på hovedområde-institution i stedet for specifikke uddannelsesudbud. En ændring i modellens fixed-effects ændrer ikke signifikansen af forældrebaggrund-variablen som prediktor af frafaldssandsynligheden. Signifikansen af de marginale effekter for forældres uddannelsesniveau er ligeledes uændret og effektstørrelserne er kun marginalt forskellige i de to modelspecifikationer. Dette er dog med undtagelse af estimationen for frafaldssandsynligheden på 2. semester. Frafaletssandsynligheden for studerende med forældre med en KVU ift. grundskole var med fixed-effects på uddannelsesudbud-niveau insignifikant (p-værdi på 0,071), men nu signifikant på et 5%-signifikansniveau (p-værdi på 0,040), se regressionstabellen i tabel 12 i appendiks A.2. På trods af den signifikante forskel mellem at have forældre med grundskole og forældre med en KVU, er

Tabel 9: Kvadreret Wald test for forældres højeste fuldførte uddannelsesniveau - model med fixed-effects på hovedområde-institution

		<i>Kvadreret Wald test, $\chi^2(5)$</i>	
	Periode for frafald	p-værdi	z test-størrelse
(1)	1. semester	0,000	24,27
(2)	2. semester	0,179	7,62
(3)	1. studieår	0,001	21,71
(4)	1. - 2. studieår	0,759	2,68
(5)	2. studieår	0,002	18,68
(6)	2. semester - 2. studieår	0,444	4,77

Note: Tallene i parentes i første kolonne refererer til perioderne i figur 1. De kvadrerede Wald test er baseret på model (10), hvor frafaldssandsynligheden er estimeret for de seks perioder.

forældres uddannelsesniveau stadig en insignifikant prediktor af frafaldssandsynligheden på 2. semester jf. den kvadreret Wald test. Hovedresultaterne fremstår robuste mod valget af uddannelsesudbud-niveaueet for modellens fixed-effects.

9.3 Kontrol for studerendes bopæl inden studiestart

I dette afsnit kontrolleres der for, om de sociale forskelle fundet i det tidlige frafald er drevet af en udeladt variabel. Det undersøges om den sociale forskel i frafaldssandsynlighed på 1. semester egentlige er udtryk for forskelle i at være tilflytter ved studiestart eller ej.

Det virker plausibelt, at studerende med højtuddannede forældre i højere grad boede tæt på universitetet inden optag, mens studerende med lavere uddannede forældre i højere grad er tilflyttere. At være tilflytter kan resultere i en mere ustabil studiestart. Det kan være gennem bekymring om ens boligsituation, følelsen af at være langt væk fra trygge, faste rammer samt ens familie og venner. Disse faktorer kan efterlade mindre fokus og overskud til de faglige og sociale aspekter ved studiestarten. Det er dokumenteret, at der eksisterer store geografiske forskelle i forældres uddannelsesniveau, hvor andelen af unge med højtuddannede forældre er størst i de større universitetsbyer og Region Hovedstaden [Høst and Sørensen, 2015]. Videre er det dokumenteret, at studerende, der har haft dårlige studiestartsoplevelser, har større frafaldssandsynlighed i løbet af det første halve år [Danmarks Evalueringsinstitut, 2017]. Som nævnt i afsnit 9.1 har studerende, som bekymrer sig meget over deres boligsituation også større frafaldssandsynlighed [Danmarks Evalueringsinstitut, 2019].

Den bopælskommune den studerende havde registreret januar samme år som studieoptag - dvs. 8-9 måneder inden studiestart - benyttes til at kontrollere for om den studerende er tilflytter. Den studerende klassificeres som tilflytter, hvis vedkommen havde bopæl i en anden kommune end universitetet, vedkommen blev optaget på. For de fire universiteter i hovedstadsområdet¹⁶, opereres der med kommuner i hovedstadsområdet¹⁷. Havde den studerende bopæl i en af disse kommuner, er vedkommen ikke tilflytter, hvis optaget på et af de fire universiteter. For RUC er Roskilde Kommune også inddraget udover

¹⁶KU, ITU, CBS og DTU.

¹⁷Kommuner i hovedstaden: København, Frederiksberg, Albertslund, Ballerup, Brøndby, Dragør, Gentofte, Gladsaxe, Glostrup, Herlev, Hvidovre, Høje Taastrup, Ishøj, Lyngby-Taarbæk, Rødovre, Tårnby, Vallensbæk, Furesø, Rudersdal, Greve.

hovedstadsområdet. Det antages, at den studerende bor tilstrækkeligt langt væk fra universitetet til, at det er nødvendigt at flytte, hvis vedkommen ikke bor i samme kommune (eller i hovedstadsområdet).

Kontrolvariablen for om den studerende er tilflytter, inkluderes i model (10), der estimeres ved MLE uden yderligere ændringer i specifikationen. Forskellen mellem de gennemsnitlige marginale effekter for forældres uddannelsesniveau i model (10) med og uden tilflytter-variablen afspejler, i hvilken grad indflydelsen fra forældrebaggrund på frafald er medieret af at være tilflytter. Regressionsresultaterne for 1. semester frafaldet er præsenteret i tabel 12 i appendiks A.2. Variablen som beskriver om den studerende er tilflytter, er insignifikant for det tidlige frafald. Betydningen af forældres uddannelsesniveau for frafaldssandsynligheden er uændret ved inddragelse af tilflytter-variablen: Signifikansniveauerne er uændrede og størrelsen af de marginale effekter er kun marginalt forskellige. Fortsat er det kun studerende med forældre med en LVU, som har signifikant lavere frafaldssandsynligheden ift. hvis forældrene har grundskole som HFU. I modellen uden tilflytter-variablen har de 2,9 procentpoint lavere frafaldssandsynlighed, og inkluderes tilflytter-variablen ændres den gennemsnitlige marginale effekt til 2,8 procentpoint.

Betydningen af forældres uddannelsesniveau for frafaldssandsynligheden på 1. semester lader ikke til at være drevet af en sammenhæng mellem forældrebaggrund og bopæl inden studiestart. Hovedresultaterne er robuste mod, om der bliver kontrolleret for, hvorvidt den studerende inden studiestart boede i samme eller omkringliggende kommune som universitetet vedkommen blev optaget på.

10 Diskussion

Der diskuteres i dette afsnit analysens data- og metodevalg samt den interne og eksterne validitet af analysens hovedresultater. Dernæst perspektiveres analysens resultater til en samfundsmæssig kontekst. Der præsenteres her hvilke konkrete politiske indsatser, som kan indføres på baggrund af resultaterne. Slutteligt kommenteres der på, hvordan videre forskning kan udbygge denne afhandlings fund.

10.1 Data- og metodiske overvejelser

Valget af grundskole som referencekategori kan diskuteres på baggrund af størrelsen af denne gruppe. Når referencekategorien til en kategorisk variabel vælges, bør den bl.a. vælges med henblik på at give et meningsfuldt sammenligningsgrundlag samt, at referencegruppen bør indeholde tilstrækkeligt mange observationer [Hardy, 1993]. Som det fremgik af den deskriptive statistik for datagrundlaget i tabel 2 udgør studerende, hvor begge forældre har grundskole som HFU 3,6 pct. af datagrundlaget svarende til 1.536 individer. En alternativ referencegruppe havde været at sammenlægge studerende med ufaglærte forældre - dvs. grundskole og gymnasial uddannelse - til én gruppe. Det havde særligt været en fordel i delanalysen i afsnit 8.2, hvor datagrundlaget opdeles i hovedområder således, at de to ufaglærte forældregrupper udgøres af relativt få observationer. Ulempen ved at sammenlægge grundskole og gymnasial uddannelse er, at det ikke er atypisk for personer i forældregenerationen med en gymnasial uddannelse at have taget en uddannelse i privat regi, som ikke registreres (eksempelvis en bankuddannelse), hvorfor de reelt ikke er ufaglærte [Heinesen, 1999]. Videre virker hovedresultaterne robuste mod ændringer i referencekategorien, hvilket er analyseret i afsnit 8.1.

I afhandlingens analyse benyttes forældres uddannelsesniveau som udtryk for de studerendes forældrebaggrund. Dette valg udfordres i robusthedsanalysen, hvor der kontrolleres for forældrenes indkomstniveau i stedet for uddannelsesniveau. Det viser sig at have en anden betydning for frafaldssandsynligheden, når forældres indkomstniveau benyttes frem for uddannelsesniveau. Alternativt til denne fremgangsmåde kunne der anvendes et sammensat indeksmål af forældres socioøkonomiske baggrund for at opfange den samlede effekt af forældrebaggrund. Dette sammensatte mål bør som minimum indeholde information om forældrenes uddannelse, indkomst og arbejdsmarkedstilknytning. Målet kan være inspireret af 'social klasser/grupper', som bl.a. er benyttet af Arbejderbevægelsens Erhvervsråd og det tidligere SFI - Det Nationale Forskningscenter for Velfærd [Pedersen and Caspersen, 2020, Ploug, 2008]. Social gruppe/klasse inddelingen har fem kategorier og baserer sig på indkomst, uddannelse samt tilknytning til arbejdsmarkedet. I OECD's PISA-undersøgelser benytter de ligeledes et sammensat indekseret mål for forældrebaggrund. PISA's 'index of economic, social and cultural status' er målt ved forældrenes uddannelsesniveau, formue og indkomst, beskæftigelse samt kulturelle ressourcer såsom antal bøger i hjemmet [OECD, 2015].

En varighedsmodel havde været en alternativ metodisk tilgang til at undersøge ændringen af betydningen af forældrebaggrund i løbet af studietiden. Mens regressionsmodeller såsom logitmodellen beskriver 'statiske' forhold, så kan varighedsmodeller benyttes til dynamiske analyser af begivenheder over tid. En varighedsmodel kan derved analysere sandsynligheden for, at et individ på et givet tidspunkt skifter fra en tilstand til en anden - eksempelvis fra at være aktiv studerende til at afbryde uddannelsen. En

varighedsmodel havde været det ideale metodevalg for denne afhandlings analyse, hvis tidspunktet for frafald var kontinuert i tid. Det ville kræve, at den konkrete dato for frafald var observeret. Som fremlagt i afsnit 5.2.1, er der stor usikkerhed forbundet med den dato, som Danmarks Statistik har registreret, som datoen uddannelsen er blevet afbrudt. Varighedsanalyse er derfor fravalgt, da den præcise dato for hvornår frafaldet indtræffer ikke er observeret.

10.2 Validitet af hovedresultater

Hovedresultaternes interne validitet hviler på, hvorvidt det kan udelukkes, at der eksisterer en uobserveret variabel, som påvirker frafaldssandsynligheden og varierer systematisk med forældrenes uddannelsesniveau. Hvis dette er tilfældet, har hovedresultaterne en lav intern validitet. Denne risiko er ved modelspecifikationen forsøgt minimeret. Der kontrolleres for en række relevante individuelle- og uddannelsesmæssige karakteristika, samt der indarbejdes fixed-effects på de specifikke uddannelsesudbud, for at opfange heterogenitet i studentergrupperne på tværs af uddannelserne. Formålet er så vidt muligt at isolere effekten fra forældrenes uddannelsesniveau på frafaldssandsynligheden. Lykkes dette, er der tale om en kausal effekt af forældres uddannelsesniveau på frafaldssandsynligheden. Det kan dog ikke udelukkes helt, at forskellen i frafaldssandsynligheden også er delvist udtryk for andre bagvedliggende forskelle mellem studerende med akademikerforældre og de studerende, hvis forældre har et lavere uddannelsesniveau.

Der argumenteres for, at analysens resultater har en høj ekstern validitet med udgangspunkt i den store population analysen er baseret på. Det forstærker den ekstern validitet, at hovedresultaterne er udarbejdet på baggrund af to års studieoptag. Det havde øget risikoen for, at resultaterne var påvirket af tidsspecifikke udsving omkring forhold, som påvirker frafald, hvis analysens datagrundlag var et enkelt års studieoptag. Det kunne give resultater, som ikke kunne generaliseres grundet eksempelvis institutionelle forhold eller politiske reformer, som særligt påvirker de studerende med optag det specifikke studieår. Det er plausibelt, at implementeringen af Studiefremdriftsreformen påvirkede studerendes frafaldssandsynlighed og muligvis den sociale skævhed heri. Reformen skærpede bl.a. studieaktivitetskravet og antal tilladte eksamensforsøg. Reformen var gældende for bachelorstuderende optaget sommeren 2014 og for øvrige universitetsstuderende fra september 2015. Studieaktivitetskravet og tvangstilmelding til eksamen blev lempet for studerende fra september 2016 og frem. Det havde været af stor bekymring for resultaternes eksterne validitet, hvis analysen havde taget udgangspunkt i studenterpopulationen med optag i 2014 eller 2015. Disse studerende var i løbet af første og andet studieår tvangstilmeldt fuldt antal ECTS-point uden mulighed for at udskyde kurser eller eksaminer, hvilket ikke er tilfældet for studerende med optag i 2016 eller senere. Med udgangspunkt i studerende optaget i 2016 og 2017, argumenteres der for, at disse studerende oplever samme institutionelle vilkår som nuværende og fremtidige bachelorstuderende (indtil nye studiereformer vedtages). Videre argumenteres der for høj ekstern validitet af analysens resultater, da datagrundlaget ikke er afgrænset til specifikke universiteter eller hovedområder, men inkluderer alle akademiske bacheloruddannelser. Det er vigtigt her at pointere, at afhandlingens resultater kun er repræsentative for de akademiske bacheloruddannelser, og ikke kan udvides til resten af det videregående uddannelsessystem uden yderligere analyse.

10.3 Politiske anbefalinger

Denne analyses hovedresultater antyder, at der eksisterer en social skævhed i frafaldssandsynligheden på 1. semester på universiteternes bacheloruddannelser. Her har de studerende med akademikerforældre signifikant lavere frafaldssandsynlighed end alle andre studerende. Forskellen i frafaldssandsynligheden er særligt udtalt mellem de studerende, hvis forældre har henholdsvis grundskole og LVU som HFU. De studerende hvis forældre har grundskole som HFU har - i gennemsnit på tværs af alle individer i datagrundlaget - 30 pct. højere frafaldssandsynligheden på 1. semester end studerende med akademikerforældre. På trods af mange års politiske indsatser med de bedste intentioner, så ses der stadig et socialt skel i frafaldet på universiteterne. Selvom frafald på de danske universitetsuddannelser i høj grad tilskrives individuelle baggrundsfaktorer, kan der være en social gevinst ved politiske indsatser fokuseret på 1. semester.

Der præsenteres her to konkrete, forskningsbaserede indsatser til at mindske den sociale skævhed i frafaldssandsynligheden. Indsatserne skal være særligt fremtrædende det første halve år efter studiestart, da nærværende analyse viser, at det i denne tid, at forældrebaggrund har en betydning for frafald. Videre viser afhandlingens delanalyse af forskelle på tværs af hovedområder, at indsatserne bør intensiveres på det naturvidenskabelige-, samfundsvidenskabelige og sundhedsvidenskabelige hovedområde.

Flere udenlandske studier undersøger effektiviteten af forskellige forberedelseskurser tilknyttet universitetet i forbindelse med studiestart, hvor formålet er at mindske den sociale ulighed i frafald. Det britiske studie af Longfellow et al. [Longfellow et al., 2008] evaluerer på et program, hvor dygtige 3.års-studerende gennemfører skriveseminarer for 1.års-studerende med henblik på at hjælpe nye studerende uden akademiske færdigheder. Deres evaluering af forsøget viser, at 1. års-studerende som deltog i programmet, blev bedre til at forstå svære begreber og gavnede også ved følelsen af et trygt læringsmiljø og en reduktion i følelsen af intimidering af universitetsverden. Et andet britiske studie af Knox [Knox, 2005] undersøger effekten af et skotsk program, som er mere målrettet at styrke de studerende faglige selvtillid. Programmet blev udbudt som sommerkursus for studerende optaget på universitetet, og hjalp de studerende med at håndtere universitetets afleverings- og bedømmelseskriterier ved, at de studerende på programmet opnåede tilstrækkelige kernekompetencer. Programmet viste sig at have flere gavnelige effekter: de studerende havde større vedholdenhed, præsterede bedre og fik større selvtillid og tro på at kunne fuldføre uddannelsen.

En konkret indsats kan derfor være at oprette målrettede forberedelseskurser tilknyttet til universitetet, hvilke afholdes enten kort inden studiestart eller sideløbende med den første tid på studiet. Erfaring viser, at disse kurser hjælper studerende med bedre at afkode universitetskulturen, de faglige krav og nedsætter følelsen af intimidering.

Der præsenteres her en yderligere indsats, som er inspireret af resultaterne fundet af Smith et al. [Smith et al., 2020], der blev præsenteret i afhandlingens litteratur gennemgang, se afsnit 4. De finder, at studerende med lavtuddannede forældre oplever at få mindre støtte og feedback fra undervisere, hvilket menes at skyldes en fremhed overfor universitetet, og at de studerende fra ikke-uddannelsesvante hjem har sværere ved at afkode de akademiske omgangsformer. Dette understøttes af teorien præsenteret i afsnit 3.3. Den signifikante betydningen af modtagelsen af feedback for frafaldssandsynligheden på universitetet understøttes yderligere af Qvortrup et al. [Qvortrup et al., 2018]. Det er derfor nærliggende med en indsats på feedback-området. International forskning viser, at feedback er et af de vigtigste elementer, når man

skal støtte studerendes læring. At de studerende modtager tilstrækkelig og brugbar feedback antages derfor at fremme den faglige integration hos studerende fra ikke-akademiske hjem.

Universitetsstuderende modtager primært summative evalueringer: de studerende får udpeget deres fejl eller mangler uden fremadskuende, konstruktiv feedback om, hvad der kunne gøres anderledes [Ulriksen et al., 2018]. Et alternativ hertil er formativ feedback, hvor de studerende lærer, hvordan de fremadrettet kan forbedre deres forståelse samtidigt med, at de lærer, hvilke bedømmelseskriterier de skal have fokus på. Det kræver væsentligt flere ressourcer at give formativ feedback sammenlignet med summativ evaluering, men jf. eksisterende litteratur og nærværende analyse kan der være et stort udbytte i at indføre formativ feedback på kurserne på 1. semester.

Der er i dette afsnit præsenteret to mulige politiske indsatser til at mindske frafaldet generelt og særligt den sociale skævhed heri. Optimalt ville begge indsatser implementeres for at komplementere hinanden. Fra både afhandlingens litteratur gennemgang samt gennemgangen af Tintos frafaldsmodel fremgår, at faglig og social integration er central for at forstå frafald. Den faglige og social integration favner bredt og indsatser for at fastholde flere studerende med ikke-uddannelsesvante forældre skal rettes flere steder. Det er vigtigt at anerkende, at frafaldet på universitetsuddannelser kun til en vis grad kan påvirkes af tiltag fra institutionsniveau, da frafald i høj grad afhænger af individuelle baggrundsfaktorer.

10.4 Videre forskning

Den videre forskning bør undersøge, hvilke mekanismer, der driver den sociale skævhed i frafaldssandsynligheden på 1. semester. Det vil kunne skabe mere målrettede politiske indsatser, hvis mekanismerne kendes. Smith et al. [Smith et al., 2020] har undersøgt dette ved surveydata, men deres population er begrænset til studerende på Aarhus Universitets humanistiske fakultet. Der bør derfor udføres en større empirisk undersøgelse baseret på bl.a. surveydata med fokus på, hvorfor studerende fra ikke-akademiske hjem i højere grad falder fra på universitetet. Undersøgelsen skal favne hele universitetssektoren og gerne gentages for flere studieoptag, for at undgå årsspecifikke udsving i besvarelsene.

Forskelle i forventninger til afkastet af uddannelse og derved motivationen for at fuldføre uddannelsen er en mulig mekanisme, som kan undersøges nærmere. Et studie af Boneva & Rauh undersøger socioøkonomiske forskelle i britiske gymnasieelevers perception af det finansielle og ikke-finansielle afkast af en universitetsuddannelse [Boneva and Rauh, 2017]. Deres studie viser, at studerende med lav socioøkonomisk status opfatter de finansielle og ikke-finansielle afkast af en universitetsuddannelse, som værende signifikant lavere end studerende med høj socioøkonomisk status opfatter dem. Den videre forskning kan undersøge, om der ses lignende forskelle i perceptionen af afkastet af en universitetsuddannelse hos danske studerende. Hvis det er tilfældet, er det plausibelt, at studerende med lavere socioøkonomisk status er mindre motiveret til at gennemføre den universitetsuddannelse de er optaget på.

11 Konklusion

Et socialt afbalanceret uddannelsessystem er et af de vigtigste instrumenter til at bryde den sociale arv. Der findes evidens for, at den sociale mobilitet i Danmark er blevet forringet i løbet af det seneste årti. Det er nødvendigt med fornyet politisk fokus på den eksisterende sociale skævhed i uddannelsessystemet. I nærværende afhandling undersøges betydningen af den studerendes forældrebaggrund for frafaldssandsynligheden i løbet af de to første studieår. Der fokuseres på, hvorvidt betydningen af forældrebaggrund ændres i løbet af studietiden. Analysen tager udgangspunkt i studerende på de akademiske bacheloruddannelser, da den sociale skævhed i uddannelsessystemet er mest udtalt i universitetssektoren.

Til at isolere effekten af forældrebaggrund på frafaldssandsynligheden benyttes en logitmodel, som kontrollerer for en række relevante baggrundskarakteristika og den uobserverede heterogenitet i frafaldssandsynligheden på tværs af uddannelsesudbud. For at undersøge ændringen i løbet af studietiden, inddeles de første to studieår i seks forskellige perioder afgrænset ved semesterskift. Frafaldssandsynligheden estimeres ved maximum-likelihood estimation for de seks forskellige perioder.

Den økonometriske analyses hovedresultater viser, at der eksisterer en social skævhed i frafaldssandsynligheden på 1. semester på universiteternes bacheloruddannelser. Estimeres frafaldssandsynligheden for perioder, hvor det første halve år ikke er indeholdt, spiller forældrenes uddannelsesniveau ingen rolle. På 1. semester har studerende med mindst én akademikerforældre statistisk signifikant lavere frafaldssandsynlighed end alle andre studerende. De bachelorstuderende hvis forældre har grundskole som deres højeste fuldførte uddannelsesniveau, har 30 pct. højere sandsynligheden for frafald på 1. semester end studerende, hvis forældre har en lang videregående uddannelse. En analyse af forskelle mellem universiteternes fem hovedområder viser, at den sociale skævhed i frafaldssandsynligheden er særlig udtalt på naturvidenskab, samfundsvidenskab og sundhedsvidenskab.

Til at forstå, hvad der driver den sociale skævhed i de studerendes frafaldssandsynlighed inddrages sociologisk teori. Jf. denne sker frafald som resultat af samspillet mellem den studerendes baggrundskarakteristika og uddannelsen. Her er det plausibelt, at studerende fra ikke-akademiske hjem i højere grad oplever en negativ eller utilstrækkelig tilknytning til uddannelsen. Det kan bl.a. forklares ved, at forældre med lange uddannelser i højere grad kultiverer uddannelsesrelevante evner hos deres børn gennem familiens opdragelsesmønstre og hverdagsstruktur. Videre har studerende med akademikerforældre en stærkere motivation for selv at fuldføre en akademisk uddannelse for at eliminere risikoen for nedadgående mobilitet. En mulig forklaring på hvorfor forældrebaggrund kun har betydning for det tidlige frafald er, at studerende på baggrund af første tilbagemelding om egne evner (ved første karaktergivning), opdaterer deres vurdering af, hvorvidt deres investering i uddannelsen kan betale sig. Heri kan der eksistere en social skævhed drevet af den lavere motivation og faglige selvtillid hos studerende fra ikke-akademiske hjem.

På tværs af det danske uddannelsessystem er der et stort fokus på at mindske frafaldet. Denne afhandlings resultater understreger vigtigheden i, at uddannelsesinstitutionerne er opmærksomme på, at de studerende møder universitetsverden med forskellige sociale baggrunde, hvilket giver dem forskellige forudsætninger for at lykkes med at fuldføre uddannelsen. Det giver anledning til fremtidige politiske interventioner med henblik på at sikre en mere socialt afbalanceret universitetssektor. Indsatserne bør

være rettet mod fastholdelse af studerende fra ikke-uddannelsesvante hjem gennem tiltag koncentreret omkring studiestart og det efterfølgende halve år.

12 Litteraturliste

Litteratur

- [Aina et al., 2018] Aina, C., Baici, E., Casalone, G., and Pastore, F. (2018). The economics of university dropouts and delayed graduation: a survey. *IZA Institute of Labor Economics*.
- [Andersen, 2018] Andersen, T. M. (2018). Human capital, inequality, and growth. *The Welfare State Revisited*, pages 109–122.
- [Arcidiacono et al., 2016] Arcidiacono, P., Aucejo, E., Maurel, A., and Ransom, T. (2016). College attrition and the dynamics of information revelation. *NBER Working Paper (No. w22325)*.
- [Becker, 1964] Becker, G. (1964). *Human capital*, volume 2. Columbia University Press.
- [Bolvig et al., 2019] Bolvig, I., Jeppesen, T., Kleif, H. B., Østergaard, J., Iversen, A., Lips, N. B., Jensen, N. L., and Thodsen, J. (2019). Unge uden job og uddannelse – hvor mange, hvorfra, hvorhen og hvorfor? *VIVE - Det Nationale Forsknings-og Analysecenter for Velfærd*.
- [Boneva and Rauh, 2017] Boneva, T. and Rauh, C. (2017). Socio-economic gaps in university enrollment: The role of perceived pecuniary and non-pecuniary returns. *CESifo Working Paper Series*.
- [Boudon, 1974] Boudon, R. (1974). Education, opportunity, and social inequality: Changing prospects in western society.
- [Breen and Goldthorpe, 1997] Breen, R. and Goldthorpe, J. H. (1997). Explaining educational differentials: Towards a formal rational action theory. *Rationality and society*, 9(3):275–305.
- [Carneiro et al., 2003] Carneiro, P., Hansen, K. T., and Heckman, J. J. (2003). Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college. Technical report, National Bureau of Economic Research.
- [Cunha et al., 2005] Cunha, F., Heckman, J., and Navarro, S. (2005). Separating uncertainty from heterogeneity in life cycle earnings. *oxford Economic papers*, 57(2):191–261.
- [Danmarks Evalueringsinstitut, 2015] Danmarks Evalueringsinstitut (2015). *Den sociale profil i optage-systemet*. København.
- [Danmarks Evalueringsinstitut, 2017] Danmarks Evalueringsinstitut (2017). *Studiestartens betydning for frafald på videregående uddannelser*. København.
- [Danmarks Evalueringsinstitut, 2018] Danmarks Evalueringsinstitut (2018). *Gymnasiale karakterers betydning for gennemførelstid på universiteterne*. København.
- [Danmarks Evalueringsinstitut, 2019] Danmarks Evalueringsinstitut (2019). *Studerendes boligsituation og frafald*. København.
- [Danmarks Evalueringsinstitut, 2020] Danmarks Evalueringsinstitut (2020). *Sociale forskelle i studerendes uddannelsesforventninger*. København.

- [Danske Universiteter, 2020] Danske Universiteter (2020). *Tal om danske universiteter 2019*. København.
- [DEA, 2019] DEA (2019). Tendenser for social mobilitet og uddannelse.
- [DEA, 2020] DEA (2020). Social mobilitet på de videregående uddannelser.
- [Hardy, 1993] Hardy, M. A. (1993). *Regression with dummy variables*. Number 93. Sage.
- [Heinesen, 1999] Heinesen, E. (1999). Den sociale arvs betydning for unges valg og resultater i uddannelsessystemet. *SFI - Socialforskningsinstituttet*.
- [Holm and Jæger, 2008] Holm, A. and Jæger, M. M. (2008). Does relative risk aversion explain educational inequality? a dynamic choice approach. *Research in Social Stratification and Mobility*, 26(3):199–219.
- [Høst and Sørensen, 2015] Høst, A. K. and Sørensen, E. S. (2015). Unges afstand til de videregående uddannelser. *Kraks Fond Byforskning*.
- [Karlson, 2014] Karlson, K. B. (2014). Præferencer for reproduktion? en analyse af den sociale stratificering af danske unges uddannelsesforventninger i henholdsvis 1968 og 2011. *Dansk Sociologi*, 25(3):93–114.
- [Karlson and Landersø, 2020] Karlson, K. B. and Landersø, R. (2020). The making and unmaking of educational opportunity: Intergenerational educational mobility in 20th century-denmark.
- [Knox, 2005] Knox, H. (2005). Making the transition from further to higher education: the impact of a preparatory module on retention, progression and performance. *Journal of Further and Higher Education*, 29(2):103–110.
- [Larsen et al., 2013] Larsen, M. R., Sommersel, H. B., and Larsen, M. S. (2013). *Evidence on dropout phenomena at universities*. Danish Clearinghouse for educational research Copenhagen.
- [Lochner, 2011] Lochner, L. (2011). Nonproduction benefits of education. In *Handbook of the Economics of Education*, volume 4, pages 183–282. Elsevier.
- [Longfellow et al., 2008] Longfellow, E., May, S., Burke, L., and Marks-Maran, D. (2008). ‘they had a way of helping that actually helped’: a case study of a peer-assisted learning scheme. *Teaching in Higher Education*, 13(1):93–105.
- [Mattsson and Munk, 2008] Mattsson, C. and Munk, M. D. (2008). Social uddannelsesmobilitet på kandidat- og forskeruddannelser. *SFI - Socialforskningsinstituttet*.
- [Nielsen et al., 2015] Nielsen, C., Munk, M. D., Jensen, M. T., Karmsteen, K., and Jørgensen, A.-M. K. (2015). Mønsterbryderindsatser på de videregående uddannelser. *SFI - Socialforskningsinstituttet*.
- [Nielsen, 2015] Nielsen, H. B. (2015). *Introduction to likelihood-based estimation and inference*. Københavns Universitet, Økonomisk Institut, København. s. 107–114.
- [OECD, 2015] OECD (2015). Pisa 2015 results (volume ii). page 205.
- [Pedersen and Caspersen, 2020] Pedersen, S. T. P. and Caspersen, S. (2020). Din klasse følger dig gennem livet. *Arbejderbevægelsens Erhvervsråd*.

- [Pihl, 2009] Pihl, M. D. (2009). Store forskelle i danskernes indkomst gennem livet. *Arbejderbevægelsens Erhvervsråd*.
- [Ploug, 2008] Ploug, N. (2008). Danskernes levkår – en solstrålehistorie. *SFI- Socialforskningsinstituttet*.
- [Qvortrup et al., 2018] Qvortrup, A., Smith, E., Rasmussen, F., and Lykkegaard, E. (2018). Studiemiljø og frafald i videregående uddannelser: Betydningen af undervisning, faglig identifikation og social integration. *Dansk Universitetspædagogisk Tidsskrift*, 13(25):151–178.
- [Roemer, 2002] Roemer, J. E. (2002). Equality of opportunity: A progress report. *Social Choice and Welfare*, 19(2):455–471.
- [Skaksen, 2018] Skaksen, J. R. (2018). *Afkast af uddannelse: Det samfundsmæssige og individuelle rationale*. Rockwool Fondens Forskningsenhed og Syddansk Universitetsforlag.
- [Smith et al., 2020] Smith, E., Reimer, D., and Kjeldsen, C. C. (2020). Betydning af individuelle faktorer for frafaldsrisiko-en analyse af spørgeskemaer og studieadministrative data. *Dansk Universitetspædagogisk Tidsskrift*, 16(28):164–181.
- [Smith and Naylor, 2001] Smith, J. P. and Naylor, R. A. (2001). Dropping out of university: a statistical analysis of the probability of withdrawal for uk university students. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 164(2):389–405.
- [Sørensen et al., 2013] Sørensen, P. B., Dalgaard, C.-J., Gjersing, A., Nikolaisen, H., Raaschou-Nielsen, A., Schröder, P., Skaksen, J. R., Sloth, B., and Sørensen, A. (2013). Uddannelse og innovation: Analyserapport 4 fra produktivitetskommissionen. *Produktivitetskommissionen*.
- [Stinebrickner and Stinebrickner, 2012] Stinebrickner, T. and Stinebrickner, R. (2012). Learning about academic ability and the college dropout decision. *Journal of Labor Economics*, 30(4):707–748.
- [Thomsen and Andrade, 2016] Thomsen, J. and Andrade, S. B. (2016). Uddannelsesmobilitet i danmark. *SFI - Socialforskningsinstituttet*.
- [Thomsen, 2017] Thomsen, J.-P. (2017). Unrealized potential: The disparate educational pathways of highly qualified working-class. *The Rockwool Foundation Research Unit*.
- [Thomsen and Bom, 2019] Thomsen, J.-P. and Bom, L. H. (2019). Danske unges uddannelsesforventninger: Med særligt fokus på unge fra kortuddannede hjem. *VIVE - Det Nationale Forsknings-og Analysecenter for Velfærd*.
- [Tinto, 1975] Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of educational research*, 45(1):89–125.
- [Troelsen, 2011] Troelsen, R. (2011). Frafall på de videregående uddannelser–hvad ved vi om årsagerne? *Dansk Universitetspædagogisk Tidsskrift*, 6(10):37–44.
- [Uddannelses- og Forskningsministeriet, 2008] Uddannelses- og Forskningsministeriet (2008). Analyse af social uddannelsesmobilitet og frafall på lange videregående uddannelser. *Idégruppen for social arv*.

- [Uddannelses- og Forskningsministeriet, 2018] Uddannelses- og Forskningsministeriet (2018). *Frafald blandt akademiske bachelorstuderende fra universiteterne inden for 4. år efter studiestart*. København.
- [Ulriksen et al., 2015] Ulriksen, L., Madsen, L. M., and Holmegaard, H. T. (2015). What make them leave and where do they go? non-completion and institutional departures in stem. *Understanding Student Participation and Choice in Science and Technology Education*.
- [Ulriksen et al., 2018] Ulriksen, L., Madsen, L. M., and Holmegaard, H. T. (2018). Indsatser mod frafald. *Institut for Naturfagenes Didaktik, Københavns Universitet*.
- [Videnskabsministeriet, 2008] Videnskabsministeriet (2008). Analyse af social uddannelsesmobilitet og frafald på lange videregående uddannelser. *Idégruppen for at bryde den sociale arv*.
- [Vignoles and Powdthavee, 2009] Vignoles, A. F. and Powdthavee, N. (2009). The socioeconomic gap in university dropouts. *The BE journal of economic analysis & policy*, 9(1).
- [Wooldridge, 2016] Wooldridge, J. M. (2016). *Introductory Econometrics*. Cengage Learning, Boston, MA, USA. s. 524–558.

A Appendiks

A.1 Regressionstabeller

Tabel 10: Regressionstabel, gennemsnitlige marginale sandsynligheder

Variable	(1)	(2)	(3)	(4)	(5)	(6)
	1. semester	2. semester	1. studieår	1. - 2. studieår	2. studieår	2. semester - 2. studieår
Gymnasial	-0,0069 (0,0128)	-0,0092 (0,0142)	-0,0161 (0,0189)	0,0148 (0,0171)	0,0014 (0,0233)	0,0074 (0,0212)
Erhvervsfaglig	-0,0074 (0,0094)	-0,0069 (0,0099)	-0,0139 (0,0135)	-0,0126 (0,0080)	-0,0171 (0,0155)	-0,0097 (0,0125)
KVU	-0,0128 (0,0113)	-0,0203* (0,0112)	-0,0330** (0,0156)	0,0024 (0,0098)	-0,0316* (0,0177)	-0,0184 (0,0145)
MVU	-0,0095 (0,0094)	-0,0070 (0,0102)	-0,0161 (0,0137)	-0,0034 (0,0080)	-0,0204 (0,0156)	-0,0107 (0,0127)
LVU	-0,0289*** (0,0095)	-0,0108 (0,0103)	-0,0392*** (0,0137)	-0,0037 (0,0082)	-0,0433*** (0,0158)	-0,0147 (0,0129)
Alder	-0,0110*** (0,0034)	-0,0053 (0,0037)	-0,0166*** (0,0048)	0,0106*** (0,0032)	-0,0052 (0,0055)	0,0062 (0,0047)
Alder ²	0,0002*** (0,0001)	0,0001 (0,0001)	0,0003*** (0,0001)	-0,00015** (0,0001)	0,0001 (0,0001)	-0,0001 (0,0001)
Kvinde	0,0051 (0,0035)	-0,0092** (0,0041)	-0,0039 (0,0051)	-0,0083** (0,0039)	-0,0115** (0,00587)	-0,0171*** (0,00528)
Ikke_vestlig	-0,0185*** (0,0061)	-0,0073 (0,0069)	-0,0259*** (0,0089)	0,0078 (0,0062)	-0,0173* (0,0102)	0,0007 (0,0088)
HF	0,0476*** (0,0051)	0,0609*** (0,0061)	0,1100*** (0,0078)	0,0391*** (0,0056)	0,1540*** (0,0094)	0,1010*** (0,0081)
HHX	0,0163*** (0,0049)	0,0044 (0,0058)	0,0204*** (0,0071)	0,0038 (0,0053)	0,0254*** (0,0081)	0,0085 (0,0072)
HTX	-0,0072 (0,0060)	-0,0051 (0,0065)	-0,0123 (0,0083)	0,0274*** (0,0059)	0,0180* (0,0093)	0,0244*** (0,0083)
Anden_adg_udd	0,0269 (0,0269)	0,0559** (0,0277)	0,0846** (0,0375)	0,0796*** (0,0181)	0,1870*** (0,0413)	0,1480*** (0,0326)
Adg_gennemsnit	-0,0085*** (0,0011)	-0,0125*** (0,0013)	-0,0213*** (0,0016)	-0,0144*** (0,0011)	-0,0368*** (0,0018)	-0,0274*** (0,0016)
Kvotel	0,0141* (0,0078)	0,0128 (0,0093)	0,0269** (0,0121)	-0,0149 (0,0139)	0,0117 (0,0161)	-0,0017 (0,0151)
Kvotel ²	-0,0141 (0,0090)	-0,0348*** (0,0095)	-0,0494*** (0,0132)	-0,0361** (0,0140)	-0,0886*** (0,0172)	-0,0713*** (0,0156)
Første_prio	-0,0168*** (0,0042)	-0,0568*** (0,0052)	-0,0756*** (0,0064)	-0,0024 (0,0043)	-0,0797*** (0,0072)	-0,0610*** (0,0063)
Wald chi2	1030,96	1568,1	2151,49	1436,55	2972,740	2298,45
Wald chi2-test (p-val)	0,000	0,000	0,000	0,000	0,000	0,000
Observationer	42.475	42.566	42.679	42.324	42.679	42.647

Note: Cluster-robuste standard fejl i (.). Fixed-effects på uddannelsesudbud. *** p<0,01, ** p<0,05, * p<0,1.

Tabel 11: Regressionstabel - hovedområde subsamples, gennemsnitlige marginale sandsynligheder

Variable	Naturvidenskab	Humaniora	Samfundsvidenskab	Teknisk videnskab	Sundhedsvidenskab (1. studieår)
Gymnasial	0,0001 (0,0352)	0,0127 (0,0242)	-0,0185 (0,0202)	0,0208 (0,0419)	-0,1330*** (0,0432)
Erhvervsfaglig	-0,0267 (0,0260)	0,0183 (0,0177)	-0,0268* (0,0149)	0,0124 (0,0276)	-0,0604* (0,0364)
KVU	-0,0565** (0,0278)	0,0307 (0,0253)	-0,0296* (0,0166)	0,0089 (0,0312)	-0,1110*** (0,0376)
MVU	-0,0213 (0,0261)	0,0133 (0,0177)	-0,0295* (0,0151)	0,0044 (0,0268)	-0,0761** (0,0374)
LVU	-0,0422 (0,0260)	-0,0075 (0,0183)	-0,0444*** (0,0153)	-0,0135 (0,0268)	-0,0876** (0,0371)
Alder	-0,0093 (0,0062)	-0,0157** (0,0076)	-0,0142*** (0,0051)	-0,0038 (0,0087)	-0,0240** (0,0116)
Alder ²	0,0002* (0,0001)	0,0002* (0,0001)	0,0002** (0,0001)	0,0001 (0,0002)	0,0005** (0,0002)
Kvinde	0,0092 (0,0090)	0,0021 (0,0079)	0,0021 (0,0051)	0,0182 (0,0119)	-0,0080 (0,0133)
Ikke_vestlig	-0,0057 (0,0144)	-0,0102 (0,0141)	-0,0348*** (0,0095)	-0,0095 (0,0174)	-0,0492*** (0,0186)
HF	0,0328*** (0,0121)	0,0386*** (0,0111)	0,0676*** (0,0076)	0,0561*** (0,0169)	0,0187 (0,0288)
HHX	-0,0081 (0,0181)	0,0351*** (0,0101)	0,0145** (0,0064)	0,0093 (0,0297)	0,0594 (0,0371)
HTX	-0,0018 (0,0098)	-0,0279 (0,0327)	0,0265*** (0,0131)	-0,0197* (0,0119)	-0,0083 (0,0171)
Anden_adg_udd	0,0542 (0,0626)	0,0125 (0,0582)	0,0220 (0,0445)	-0,0029 (0,0881)	0,1480* (0,0778)
Adg_gennemsnit	-0,0123*** (0,0022)	-0,0038 (0,0024)	-0,0082*** (0,0019)	-0,0135*** (0,0033)	-0,0021 (0,0055)
Kvote1	0,0350 (0,0235)	0,0132 (0,0142)	0,0045 (0,0134)	0,0190 (0,0305)	0,0245 (0,0218)
Kvote2	-0,0206 (0,0220)	-0,0017 (0,0205)	-0,0228 (0,0144)	-0,0035 (0,0316)	-0,0105 (0,0262)
Første_prio	-0,0348*** (0,0099)	-0,0130 (0,0096)	-0,0079 (0,0058)	-0,0352*** (0,0134)	-0,0969*** (0,0124)
Wald chi2	228,00	230,96	1030,96	158,62	275,07
Wald chi2-test (p-val)	0,000	0,000	0,000	0,000	0,000
Observationer	7.711	11.231	15.715	4.417	3.401

Note: Cluster-robuste standard fejl i (.). Fixed-effects på uddannelsesudbud. Estimeret for 1. semester og 1. studieår for sundhedsvidenskab. *** p<0,01, ** p<0,05, * p<0,1.

Table 12: Regressionstabel - tilflytter-variabel og fixed-effects på hovedområde-institution, gennemsnitlige marginale sandsynligheder

Variable	(1a)		(1)		(2)		(3)		(4)		(5)		(6)	
	1. semester	2. semester	1. semester	2. semester	1. semester	2. semester	1. studieår	2. studieår	1. - 2. studieår	1. studieår	2. studieår	1. - 2. studieår	2. semester - 2. studieår	2. studieår
Gymnasial	-0,0063 (0,0128)	-0,0047 (0,0134)	-0,0092 (0,0197)	0,0250 (0,0245)	0,0171 (0,0300)	0,0213 (0,0283)								
Erhvervsfaglig	-0,0061 (0,0093)	-0,0087 (0,0100)	-0,0149 (0,0149)	0,0033 (0,0082)	-0,0187 (0,0175)	-0,0098 (0,0137)								
KVU	-0,0120 (0,0113)	-0,0148 (0,0125)	-0,0392** (0,0118)	0,0028 (0,0106)	-0,0372* (0,0201)	-0,0220 (0,0157)								
MVU	-0,0089 (0,00936)	-0,0101 (0,0098)	-0,0183 (0,0151)	-0,0039 (0,0082)	-0,0229 (0,0175)	-0,0124 (0,0138)								
LVU	-0,0284*** (0,0095)	-0,0287*** (0,0100)	-0,0406*** (0,0152)	-0,0052 (0,0083)	-0,0457** (0,0178)	-0,0174 (0,0140)								
Tilflytter	-0,0039 (0,0035)	-	-	-	-	-								
Alder	-0,0118*** (0,0034)	-0,0121*** (0,0036)	-0,0199*** (0,0050)	0,0094*** (0,0033)	-0,0100* (0,0058)	0,0024 (0,0050)								
Alder ²	0,0002*** (0,0001)	0,0002*** (0,0001)	0,0003*** (0,0001)	-0,0001** (0,0001)	0,0020* (0,0001)	0,0000 (0,0001)								
Kvinde	0,0047 (0,0035)	0,0019 (0,0036)	-0,0114*** (0,0052)	-0,00963** (0,0042)	-0,0203*** (0,0062)	-0,0226*** (0,0055)								
Ikke_vestlig	-0,0197*** (0,0061)	-0,0211*** (0,0062)	-0,0291*** (0,0093)	0,0079 (0,0077)	-0,0197* (0,0119)	0,0005 (0,0104)								
HF	0,0474*** (0,0051)	0,0499*** (0,0052)	0,1140*** (0,0080)	0,0422*** (0,0060)	0,1610*** (0,0099)	0,1050*** (0,0085)								
HHX	0,0165*** (0,0050)	0,0186*** (0,0052)	0,0267*** (0,0058)	-0,0058 (0,0054)	0,0232*** (0,0088)	0,0037 (0,0076)								
HTX	-0,0072 (0,0060)	-0,0061 (0,0060)	-0,0102 (0,0086)	0,0317*** (0,0066)	0,0257** (0,010)	0,0308*** (0,0091)								
Anden_adg_udd	0,0197 (0,0280)	0,0236 (0,0275)	0,0839** (0,0366)	0,0802*** (0,0184)	0,1920*** (0,0396)	0,1520*** (0,0313)								
Adg_gennemsnit	-0,0087*** (0,0011)	-0,0102*** (0,0010)	-0,0230*** (0,0015)	-0,0145*** (0,0011)	-0,0384*** (0,0018)	-0,0276*** (0,0016)								
Kvotet1	0,0146* (0,0078)	0,0106** (0,0047)	0,0100 (0,0076)	-0,0005 (0,0068)	0,0075 (0,0096)	-0,0011 (0,0088)								
Kvotet2	-0,0140 (0,0090)	-0,0232*** (0,0069)	-0,0749*** (0,0093)	-0,0246*** (0,0076)	-0,1040*** (0,0116)	-0,0764*** (0,0098)								
Første_prio	-0,0164*** (0,0042)	-0,0180*** (0,0042)	-0,0795*** (0,0070)	-0,0051 (0,0044)	-0,0861*** (0,0079)	-0,0662*** (0,0072)								
Wald chi2	1026,95	506,21	1156,19	830,44	1699,06	1372,25								
Wald chi2-test (p-val)	0,000	0,000	0,000	0,000	0,000	0,000								
Observationer	42.030	42.681	42.681	42.681	42.681	42.681								

Note: Cluster-robuste standard fejl i (.). Fixed-effects på uddannelsesudbud i kolonne (1a) og på hovedområde-institution i kolonne (1)-(6). *** p<0,01, ** p<0,05, * p<0,1.

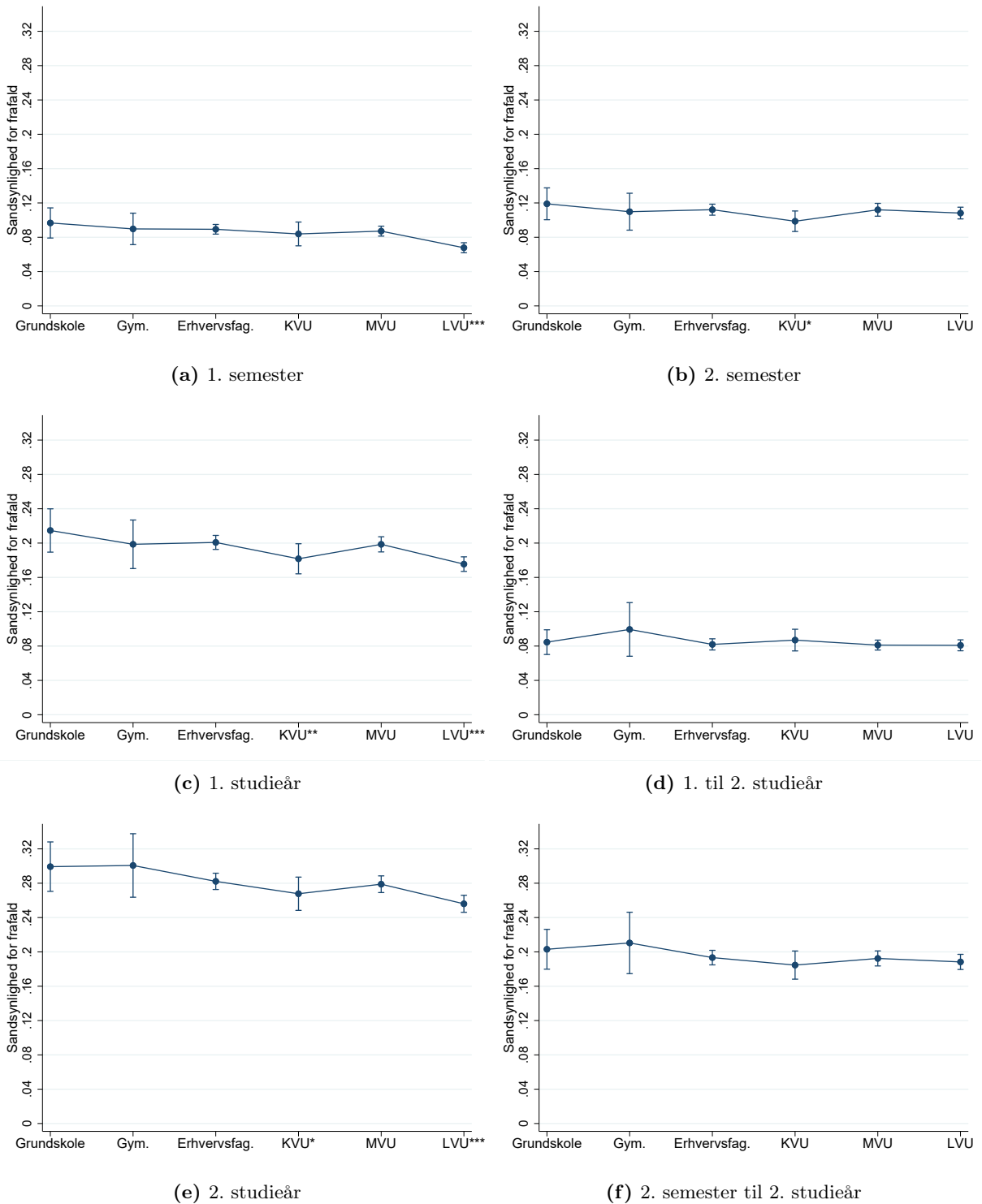
Tabel 13: Regressionsstabel - forældres indkomstniveau, gennemsnitlige marginale sandsynligheder

Variable	(1)	(2)	(3)	(4)	(5)	(6)
	1. semester	2. semester	1. studieår	1. - 2. studieår	2. studieår	2. semester - 2. studieår
Mor_indkomst (log)	-0,0019 (0,0026)	-0,0002 (0,0032)	-0,0020 (0,0040)	-0,0019 (0,0028)	-0,0039 (0,0046)	-0,0022 (0,0040)
Far_indkomst (log)	-0,0018 (0,0017)	-0,0032 (0,0025)	-0,0050 (0,0031)	-0,0044** (0,0021)	-0,0093** (0,0037)	-0,0076** (0,0033)
Alder	-0,0124*** (0,0035)	-0,0030 (0,0038)	-0,0157*** (0,0050)	0,0107*** (0,0034)	-0,0043 (0,0056)	0,0084* (0,0048)
Alder ²	0,0002*** (0,0001)	0,0000 (0,0001)	0,0003*** (0,0001)	-0,0002** (0,0001)	0,0001 (0,0001)	-0,0001 (0,0001)
Kvinde	0,00683* (0,0036)	-0,0102** (0,0042)	-0,0032 (0,0052)	-0,0090** (0,0040)	-0,0114* (0,0059)	-0,0188*** (0,0053)
Ikke_vestlig	-0,0173*** (0,0062)	-0,0089 (0,0070)	-0,0262*** (0,0090)	0,0040 (0,0064)	-0,0216** (0,0103)	-0,0049 (0,0090)
HF	0,0507*** (0,0051)	0,0605*** (0,0061)	0,1130*** (0,0078)	0,0393*** (0,0056)	0,1570*** (0,0095)	0,1010*** (0,0081)
HXX	0,0194*** (0,0049)	0,0049 (0,0058)	0,0240*** (0,0071)	0,0044 (0,0053)	0,0294*** (0,0081)	0,0095 (0,0073)
HTX	-0,0053 (0,0061)	-0,0064 (0,0065)	-0,0117 (0,0084)	0,0275*** (0,0059)	0,0191** (0,0094)	0,0235*** (0,0083)
Anden_adg_udd	0,0361 (0,0268)	0,0631** (0,0271)	0,1000*** (0,0363)	0,0786*** (0,0182)	0,2010*** (0,0404)	0,1530*** (0,0325)
Adg_gennemsnit	-0,0093*** (0,0011)	-0,0123*** (0,0013)	-0,0219*** (0,0016)	-0,0144*** (0,0011)	-0,0374*** (0,0018)	-0,0273*** (0,0016)
Kvot1	0,0138* (0,0080)	0,0147 (0,0094)	0,0285** (0,0123)	-0,0143 (0,0144)	0,0141 (0,0164)	0,0011 (0,0154)
Kvot2	-0,0147 (0,0092)	-0,0337*** (0,0096)	-0,0489*** (0,0133)	-0,0349** (0,0147)	-0,0865*** (0,0175)	-0,0685*** (0,0159)
Første_prio	-0,0174*** (0,0042)	-0,0574*** (0,0053)	-0,0768*** (0,0065)	-0,0026 (0,0043)	-0,0811*** (0,0073)	-0,0618*** (0,0064)
Wald chi2	972,84	1555,44	2097,14	1423,28	2911,70	2264,32
Wald chi2-test (p-val)	0,000	0,000	0,000	0,000	0,000	0,000
Observationer	41.649	41.740	41.851	41.462	41.851	41.819

Note: Cluster-robuste standard fejl i (.). Fixed-effects på uddannelsesudbud. *** p<0,01, ** p<0,05, * p<0,1.

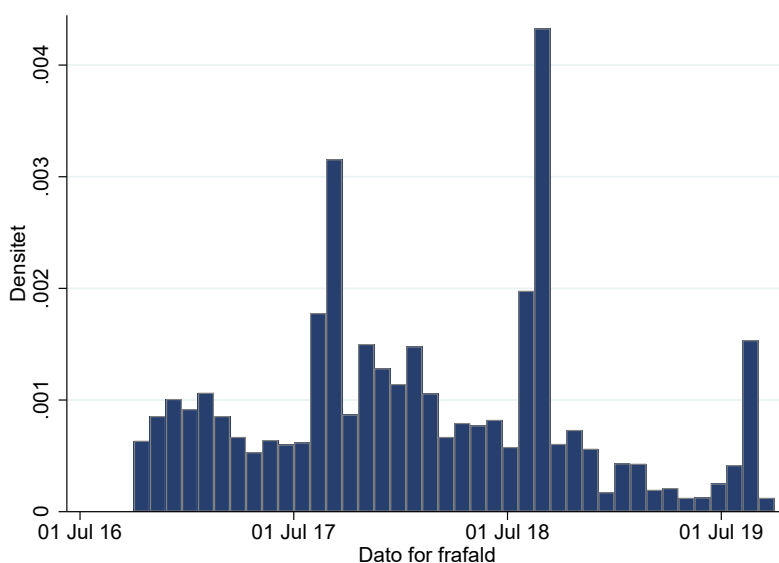
A.2 Diverse

Figur 8: Prædikterede marginale sandsynligheder uden referencekategori med 95% konfidensinterval



Note: X-akserne viser forældres højeste fuldførte uddannelsesnivea. Frafallssandsynligheden er estimeret ved model (10) for de seks perioder præsenteret i figur 1. Der refereres til signifikansniveauet af forældres uddannelsesnivea relativt til hvis forældrene havde grundskole som højeste fuldførte uddannelsesnivea: *** $p < 0,01$, ** $p < 0,05$, * $p < 0,1$.

Figur 9: Histogram for frafaldsdatoen



Note: Dato for frafald er den dato Danmarks Statistik har registreret, at den påbegyndte uddannelse er blevet afbrudt. Baseret på de studerende i datagrundlaget som er frafaldet i løbet af de to første studieår, $N=11.732$.

Tabel 14: Fordelingen i forældres højeste fuldførte uddannelsesniveau blandt optag fordelt på hovedområder

<i>Antal optaget</i>	<i>Forældres højeste fuldførte uddannelsesniveau</i>						Total
	Grundskole	Gym.	Erhvervsfag.	KVU	MVU	LVU	
Naturvidenskab	254	191	2.123	450	2.532	2.163	7.713
Teknisk videnskab	128	85	1.103	294	1.503	1.349	4.462
Humaniora	460	296	3.437	711	3.980	2.452	11.336
Samfundsvidenskab	546	403	4.650	1.008	4.891	4.271	15.769
Sundhedsvidenskab	131	141	804	230	1.089	1.006	3.401

Tabel 15: Kvadreret Wald test af mors indkomstniveau (log-transformeret)

	Periode for frafald	<i>Kvadreret Wald test, $\chi^2(1)$</i>	
		p-værdi	z test-størrelse
(1)	1. semester	0,541	0,37
(2)	2. semester	0,928	0,01
(3)	1. studieår	0,776	0,08
(4)	1. - 2. studieår	0,746	0,10
(5)	2. studieår	0,651	0,20
(6)	2. semester - 2. studieår	0,860	0,03

Note: Tallene i parentes i første kolonne refererer til perioderne i figur 1. De kvadrerede Wald test er baseret på model (10), hvor frafaldssandsynligheden er estimeret for de seks perioder.

Tabel 16: Kvadreret Wald test af fars indkomstniveau (log-transformeret)

		<i>Kvadreret Wald test, $\chi^2(1)$</i>	
	Periode for frafald	p-værdi	z test-størrelse
(1)	1. semester	0,300	1,07
(2)	2. semester	0,120	1,65
(3)	1. studieår	0,100	2,71
(4)	1. - 2. studieår	0,037	4,33
(5)	2. studieår	0,012	6,33
(6)	2. semester - 2. studieår	0,022	5,27

Note: Tallene i parentes i første kolonne refererer til perioderne i figur 1. De kvadrerede Wald test er baseret på model (10), hvor frafaldssandsynligheden er estimeret for de seks perioder.