

Anna Varpen Helgheim & Sandra Helen Iversen

Arbeid ved siden av studiene: Hvordan påvirker dette utdanningsutfallet?

Bacheloroppgave i Samfunnsøkonomi

Veileder: Bjarne Strøm

Mai 2022

Anna Varpen Helgheim & Sandra Helen Iversen

Arbeid ved siden av studiene: Hvordan påvirker dette utdanningsutfallet?

Bacheloroppgave i Samfunnsøkonomi

Veileder: Bjarne Strøm

Mai 2022

Norges teknisk-naturvitenskapelige universitet

Fakultet for økonomi

Institutt for samfunnsøkonomi



Kunnskap for en bedre verden

Sammendrag

Frafall i høyere utdanning er et viktig tema i politikken i Norge. Dette er fordi Norge genererer økonomisk vekst gjennom en høyt utdannet befolkning. I denne bacheloroppgaven er det forsøkt å kartlegge i hvilken grad frafall i høyere utdanning blir påvirket av om studentene har lønnet arbeid ved siden av studiene. For å analysere denne problemstillingen er det brukt minste kvadraters metode for en lineær sannsynlighetsmodell. Datasettet som er brukt i denne studien er hentet fra *Den europeiske studentundersøkelsen - Eurostudent VI* fra 2016. Studien viser at arbeidstid har ingen signifikant effekt på utdanningsutfallet. Samtidig viser det seg at dersom studenten jobber over en viss terskel vil dette gi en betydelig negativ effekt på utdanningsutfallet. På en annen side er det de sosiodemografiske kjennetegnene til individet som vil ha en avgjørende effekt på om studenten fortsetter påbegynt studie eller ikke. Dette samsvarer også med tidligere forskning. Til tross for dette, gir studiestøtten insentiver til å ta utdanning og sørge for at alle skal ha like muligheter til høyere utdanning.

Abstract

Dropout rates in higher education is an important topic in Norwegian politics. This is because Norway generates economic growth through a highly educated population. In this bachelor thesis, it is attempted to look at to which extent dropouts in higher education are affected by whether the students have paid work in addition to their studies. To analyze this problem, the least squared method has been used for a linear probability model. The dataset used in this study is taken from *Den europeiske studentundersøkelsen - Eurostudent VI* from 2016. The study shows that working hours have no significant effect on the educational outcome. On the other hand, it is the sociodemographic characteristics of the individual that will have a determining effect on whether the student continues their studies or not. This is also consistent with previous research. Despite this, educational grants provide incentives to take education and ensures that everyone has equal opportunities for higher education

Innholdsfortegnelse

SAMMENDRAG	1
ABSTRACT	2
1 INTRODUKSJON	4
1.1 MOTIVASJON	4
1.2 PROBLEMSTILLING.....	4
2 TEORETISK RAMMEVERK OG TIDLIGERE LITTERATUR.....	5
2.1 INNLEDNING.....	5
2.2 STUDENTENES VALG OM Å STUDERE SOM RASJONELLE INDIVID	5
2.3 SAMFUNNSØKONOMISK TEORI.....	6
2.4 FAKTORER SOM KAN PÅVIRKE FRAFALL FRA STUDIER	7
2.5 OPPSUMMERING	8
3 EMPIRISK STRATEGI	8
3.1 INNLEDNING.....	8
3.2 MULTIPPEL REGRESJONSANALYSE	9
3.2.1 Innledning	9
3.2.2 Minste kvadraters metode (OLS).....	9
3.2.3 Modellens forklaringsgrad.....	9
3.2.4 Forutsetninger for OLS	10
3.2.5 Lineær sannsynlighetsmodell	14
3.3 HYPOTESETESTING	14
3.3.2 Breusch-Pagan test	16
3.4 OPPSUMMERING	16
4 PRESENTASJON AV DATAMATERIALE	16
4.1 INNLEDNING.....	16
4.2 DATAKILDER	17
4.3 DESKRIPTIV STATISTIKK FOR DATAMATERIALET	17
4.3.1 Deskriptiv statistikk, avhengig variabel.....	18
4.3.2 Deskriptiv statistikk for interessevariabel	19
4.3.3 Deskriptiv statistikk av kontrollvariabler	20
4.5 OPPSUMMERING	24
5 EMPIRISKE RESULTATER	25
5.1 INNLEDNING.....	25
5.2 RESULTATER	25
5.2.1 Modell i: enkel lineær modell.....	25
5.2.2 Modell ii: utvidet lineær modell med kontroll for sosiodemografiske variabler	27
5.2.3 Modell iii: utvidet modell med kontroll for utdanningsvariabler.....	29
5.2.4 Modell iv: utvidet ikke-lineær modell.....	30
5.3 DISKUSJON.....	31
5.4 MULIGE UTVIDELSER	35
5.5 OPPSUMMERING	36
6 KONKLUSJON	36
7 REFERANSER.....	37
Vedlegg 1: Korrelasjonsmatrise	39

1 Introduksjon

1.1 Motivasjon

Utdanning i Norge er gratis og er ansett som et velferdsgode. Fra et makroperspektiv er utdanning en investering i humankapital som skal bidra til produktivitet og økonomisk vekst. Den skal også være med å sørge for finansiering av velferdsstaten og på den måten bidra til å redusere sosiale og økonomiske forskjeller i samfunnet. For at det skal være lønnsomt for privatpersoner å studere er det viktig at staten bidrar økonomisk og på den måten redusere alternativkostnaden. Dette er den kostnaden individene opplever av mistet inntekt i den forstand at den enkelte student går glipp av inntekt når de går på skole, i tillegg til at det tar lengre tid før de begynner å tjene penger (St.meld.13(2018-2019)).

Norge investerer store summer på høyere utdanning. Bakgrunnen for dette er at Norge er en kunnskapsnasjon som har et stort behov for personer med høy faglig kompetanse.

Ifølge OECD (2019) er den direkte kostnaden, inkludert studiestøtten, den største andelen av totale offentlige kostnader for høyere utdanning (OECD, 2019). De høye utgiftene baserer seg på at Norges investering i utdanning vil føre til en høy samfunnsøkonomisk avkastning i form av økonomisk vekst.

Studiestøtten skal i all hovedsak sørge for at alle skal ha like muligheter til å ta utdanning uavhengig av deres sosiodemografiske og økonomiske forhold. På den måten vil støtten sikre arbeidslivet tilgang til kompetente arbeidstakere (utdanningsstøtteloven, 2005, §1-1). Dermed vil den bidra med å avlaste studentenes kostnader slik at det blir lettere å fokusere på studiene. Likevel er det ikke uvanlig at studenter jobber ved siden av studiene. Et argument for å jobbe ved siden av studiene er at det kan være fordelaktig å ha arbeidserfaring når man skal søke jobb. Samtidig er det rimelig å anta, at for mye lønnet arbeid ved siden av studiene kan gi negative konsekvenser for studentene i form av mistrivsel og frafall fra studiene.

I denne oppgaven skal vi studere om det er noen sammenheng mellom studentenes frafall fra studie og lønnet arbeid.

1.2 Problemstilling

Sammenhengen mellom lønnet arbeid ved siden av studiene og utdanningsutfallet, som ble nevnt innledningsvis, motiverte oss til å undersøke hvordan arbeid under studietiden kan

påvirker studentenes valg om å fortsette studiene sine eller ikke. Derfor kom vi frem til følgende problemstilling:

“Er det sammenheng mellom studentenes frafall fra studier og lønnet arbeid?”

Vi starter med å undersøke hvordan arbeidstid og det å ha jobb vil påvirke studentenes planer om å fortsette studiene med en regresjonsmodell. Deretter utvides analysen for å studere hvordan effekten endres når det legges til kontrollvariabler. Det kontrolleres først for individets sosiodemografiske bakgrunn og deretter for utdanningsbakgrunn, studieprogram og fagfelt. Videre studerer vi om det finnes noen ikke-lineære effekter, og om det er et vendepunkt hvor mengden arbeidstid blir så høy at de slutter.

Oppgaven er organisert som følger: Kapittel 2 presenterer teoretisk rammeverk og tidligere litteratur. Kapittel 3 presenterer empirisk strategi, og kapittel 4 presenterer datamaterialet. Kapittel 5 inneholder empiriske resultater og diskusjon, mens kapittel 6 inneholder hovedfunnene og en konklusjon av analysen vår.

2 Teoretisk rammeverk og tidligere litteratur

2.1 Innledning

Studentenes valg om å ta utdanning blir påvirket av flere faktorer som blant annet sosiale og økonomiske perspektiver. Likevel er det 1 av 4 som dropper ut av studiet sitt (Andresen & Lervåg, 2022). Politikere har den siste tiden vært opptatt av å senke frafallstallet, både på høyere utdanning og videregående skolenivå. Studiestøtten er et slikt insentiv som skal hjelpe studentene økonomisk. Likevel jobber politikere med flere økonomiske insentiv som skal bidra til å senke frafallstallet. Eksempelvis, for at en skal få fullt stipend må studenten fullføre graden de har begynt på (Forskrift om utdanningsstøtte, 2020). Ved å senke frafallstallet kan dette bidra til å øke samfunnsøkonomisk avkastning.

2.2 Studentenes valg om å studere som rasjonelle individ

I en forskningsartikkel fra Jæger (2007) presenteres det en teori om sosiologisk rasjonelle valg (SRC) som baserer seg på beslutninger i ulike teoretiske perspektiver. Fra et økonomisk perspektiv vil den totale nytten av utdanningsvalg være et resultat av økonomisk og sosial

avkastning. I følge SRC-modellen er studenter rasjonelle og nyttemaksimerende individer (Jæger, 2007). En annen dimensjon er at studentene sine forventninger til avkastning av utdanning er påvirket av foreldres utdanningsbakgrunn og interaksjoner med jevnaldrende. Dette innebærer at studentenes valg av utdanning har som formål å maksimere et komplekst nytteproblem som innebærer både økonomisk og sosial avkastning til utdanning. Konsekvensen er at når studenter skal inngå et valg om deres utdanning, søker de ikke kun etter å maksimere sin nytte for en langsiktig økonomisk gevinst, men også for å øke sannsynligheten for å bevare eksisterende relasjoner til jevnaldrende.

Den totale nytten av utdanningsvalg består av tre komponenter: økonomisk avkastning, sosial avkastning og mulighetsbegrensning. Vi antar en additiv modell og får følgende nyttefunksjon

$$U_i^j = \beta_1 e_i^j + \beta_2 s_i^j + \beta_3 x_i^j + v_i^j; j = (1, \dots, J),$$

hvor den totale nytten av utdanningsvalg j ($j = 1, \dots, J$) for individ i , U_i^j , omfatter den forventede økonomiske avkastningen e , sosiale avkastningene s , og x som subsumerer mulighetsbegrensinger pålagt av sosial bakgrunn. β ($\beta = 1, 2, 3$) er parametere eller vektor av parametere som fanger vekten av hvert enkelt element i nyttefunksjonen, og v summerer nytte-genererende faktorer som er uobserverbare. Denne funksjonen tar hensyn til at familiære bakgrunn vil være med å påvirke forventede avkastning til utdanning.

2.3 Samfunnsøkonomisk teori

I en forskningsartikkel utredet av Erling (Brandt, 2005) ser på avkastning av investering i høyere utdanning, hvor det skilles mellom privatøkonomisk og samfunnsøkonomisk avkastning av utdanning. Privatøkonomisk avkastning karakteriseres av kostnader og gevinster for den enkelte student. Hovedkostnaden er ofte manglende inntekt under studietiden, mens gevinsten vil være større muligheter for høy inntekt etter fullført utdanning. Likevel, som nevnt i kapittel 2.2, er det viktig å nevne at lønn ikke alene er en avgjørende faktor for å ta høyere utdanning. Når vi inkluderer for finansiell støtte og skatt av inntekt vil dette påvirke både risikoen ved kostnader og gevinsten av høyere utdanning. Finansiell støtte som stipend og studielån reduserer inntektstapet under utdanningen, mens tilbakebetaling av studielån og høyere skatt som følge av høyere inntekt vil redusere gevinsten etter endt utdannelse. Dette kan indikere en lavere risiko ved å velge høyere utdanning, noe som igjen

kan påvirke individers studieprogresjon. Dette kan for eksempel være i form av at terskelen for å droppe ut reduseres, og antallet individer som ikke fortsetter påbegynt studie heller velger å gjøre noe annet. Den samfunnsøkonomiske avkastningen består av mange faktorer. Blant dem er kostnader og gevinster ved utdanningssystemet som ikke dekkes av studenten selv, men også nyskaping og økonomisk vekst. Utdanning fører til mer forskning og utvikling, kunnskapsoverføring mellom bedrifter, og teknisk og organisatorisk fremgang. Artikkelen fremhever at det er viktig å merke at individer som skal velge utdanning, eller om de skal fortsette på studiet de allerede har valgt, vil i liten grad ta hensyn til alle disse effektene. Videre kommer det frem at den enkelte vil ta hensyn til den privatøkonomiske avkastningen i dette valget.

Ifølge Brandt (2005) er Norge i stand til å utdanne en langt større andel av befolkningen enn de fleste andre land. Økonomiske insentivsystemer som støtte- og finansieringsordninger bidrar til å stimulere både tilbudet og etterspørselen etter høyere utdanning. Systemet vi har her i Norge som kombinerer en sammenpresset lønnsstruktur, et progressivt skattesystem, og høy studiestøtte gjør at vi har en arbeidsstyrke med svært høyt utdanningsnivå.

2.4 Faktorer som kan påvirke frafall fra studier

Hovdhaugen (2015) hevder at det finnes to ulike motivasjoner for studenter som jobber ved siden av studiene sine: Hvorvidt studentene er *nødt* til å jobbe eller hvorvidt de *velger* å jobbe.

De studentene som jobber fordi de er nødt til å forsørge seg selv har en større sannsynlighet for å avslutte studiene (Hovdhaugen, 2015). Dette kan spesielt gjelde studenter som kommer fra hjem med dårlige økonomiske betingelser hvor foreldrene ikke har anledning til å bidra til økonomisk støtte gjennom studietiden. Uavhengig av sosialdemografiske bakgrunn så er det mange studenter som velger å jobbe til tross for at det kan påvirke deres akademiske prestasjoner. I Norge får studentene studiestøtte i form av Statens Lånekasse som gjør at en kan anta at de fleste norske studenter som jobber ved siden av studiene jobber fordi de velger det selv (Hovdhaugen, 2015).

Videre berettiger Hovdhaugen (2015) at 60-70% av studentene jobber ved siden av studiene ved utenlandske utdanningsinstitusjoner. Norske studenter får statlig hjelp fra Lånekassen i form av økonomisk støtte, i tillegg er det gratis å studere ved norske offentlige

utdanningsinstitusjoner. Likevel kommer det frem i artikkelen at det er en like stor andel norske studenter som jobber som utenlandske studenter. Dette til tross for at utenlandske studenter ikke har de samme økonomiske godene som man får i Norge. I tillegg er risikoen for å droppe ut som et resultat av å jobbe, like stor for norske og utenlandske studenter. Det er verdt å merke seg at de fleste studenter som jobber for få arbeidserfaring og en fot inn i arbeidsmarkedet. Artikkelen fremhever at det finnes en terskel for hvor mye studentene kan jobbe før risikoen for frafall øker. Dersom studentene jobber mer enn 20 timer lønnet arbeidstimer i uken, vil risikoen for å droppe ut øke. Ressurssterke studenter har mindre sannsynlighet for å droppe ut uavhengig om de jobber eller ikke. Dette indikerer at økonomisk støtte og gratis høyere utdanning kun har en liten betydning for sannsynligheten for at studentene jobber og dropper ut. Dermed kan det konkluderes med at land med gratis utdanning og land med studieavgift vil ha marginale forskjeller når det gjelder studenter som jobber og avslutter studiene (Hovdhaugen, 2015).

I en annen forskningsartikkel av Moulin et al. (2013) vises det at det finnes en terskel for antall timer studentene jobber. Dersom denne terskelen overstiges, vil man få en betydelig større risiko for å droppe ut. Denne terskelen ligger på 25 lønnet arbeidstimer i uken. Dersom studentene slutter med intensjon om å bytte til et annet studium, vil ikke antall timer med lønnet arbeid ha samme effekten. (Moulin et al., 2015).

2.5 Oppsummering

I dette kapitlet innledes teorien som ligger til grunn for vårt valg av problemstilling, og det er gitt en kort gjennomgang av ulike motivasjoner til at studenter velger å ta utdanning.

Kapitlet belyser hvorfor det er viktig for Norge å senke frafallstallet. Videre redegjøres det for teori som kan forklare hvorfor studenter velger å jobbe ved siden av studiene og hvordan risikoen av antall arbeidstimer øker sannsynligheten for frafall etter en viss terskel er nådd.

3 Empirisk strategi

3.1 Innledning

Kapitlet redegjør for teorien om minste kvadraters metode og presenterer de klassiske lineære modellforutsetningene (CLM). CLM inneholder Gauss-Markov antakelsene, i tillegg til MLR.6 som presenterer et normalfordelt feilledd. Vi skal også presenterer teorien om en

lineær sannsynlighetsmodell, da vi har en binær avhengig variabel. Til slutt vil vi introdusere teorien om hypotesetesting. Disse teoriene skal vi bruke i oppgaven vår for å analysere om det er en sammenheng mellom utdanningsutfall og jobb ved siden av studie. Siden vi ser på flere eksogene variabler som kan påvirke den endogene variabelen, skal vi hovedsakelig bruke en multipl regressjonsanalyse i vår oppgave.

3.2 Multipl regressjonsanalyse

3.2.1 Innledning

I en multipl regressjonsanalyse kan eksplisitt kontrollere for flere faktorer som simultant påvirker den avhengige variabelen, y_1 . En multipl regressjonsmodell kan bedre predikere hvordan den avhengige variabelen blir påvirket. Videre introduseres teorien om minste kvadraters metode og hvilke forutsetninger som må holde for at metoden skal kunne brukes i analysen.

3.2.2 Minste kvadraters metode (OLS)

Minste kvadraters metode («Ordinary Least Squares», forkortet som OLS) brukes til å estimere parametere i en multipl regressjonsanalyse. Ved OLS velges estimatorer for å minimere summen av kvadratiske residual (SSR), der SSR måler variabiliteten i feilledet u til regressjonsmodellen. Det betyr at gitt n observasjoner på y , x_1 , og x_2 , vil estimatene $\hat{\beta}_0$, $\hat{\beta}_1$ og $\hat{\beta}_2$ bli valgt simultant for å gjøre $\sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2})^2$ minst mulig. (Wooldridge, 2019, s.70)

Det som omtales som *utelatt variabelproblem* beskriver at det utelates en variabel som er relevant for modellen, noe som vil føre til skjevhet ved at estimerte parametere ikke blir lik populasjonsparameterne i forventningsforstand. Dette fører til en invalid eller upålitelig OLS estimator. Dette vil vi se nærmere på i kapittel 5.3 når vi tolker resultatene av analysen.

3.2.3 Modellens forklaringsgrad

R^2 er modellens forklaringsgrad. R^2 er et mål på hvor mye modellen forklarer av variasjon i y_i . Den totale variasjonen (SST) består av den forklarte variasjonen (SSE) og den uforklarte variasjonen (SSR). (Wooldridge, 2019, s.76). Modellens forklaringsgrad er gitt ved:

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}, \quad 0 < R^2 < 1 \quad (3.1)$$

Fra formel 3.1 har vi at R^2 vil øke når det legges til flere relevante uavhengige variabler i regresjonsmodellen. Det vil altså si at jo flere relevante faktorer som inkluderes i modellen, jo mer vil det kunne trekkes slutninger om variasjonen i den forklarte variabelen.

3.2.4 Forutsetninger for OLS

For at regresjonsanalysen skal være gjeldende og for at man kan bruke modellen til å trekke konklusjoner forutsettes det at Gauss-Markovs antakelsene er valide. I tillegg til MLR.1-MLR.5, presenteres teorien om MLR.6 som omhandler normalfordeling av feilledet.

MLR.1 Linearitet

Den første antakelsen definerer den multiple regresjonsmodellen, og forutsetter at modellen må være lineær i parameterne $(\beta_0, \dots, \beta_k)$ for å holde. Modellen må også inneholde et feilledd u som fanger opp eventuelle andre uobserverte faktorer som kan påvirke utfallet.

Det er økonomiske sammenhenger som åpner for at modellen kan ha fleksibel linearitet. Den avhengige og de uavhengige variablene kan være vilkårlige funksjoner av de underliggende variablene av interesse. Eksempler på dette er kvadrater, noe vi vil komme tilbake til i kapittel 5.1. Foreløpig antar vi linearitet i vår modell (Wooldridge, 2019, s.80).

MLR.2 Tilfeldig utvalg

Denne antakelsen forutsetter at datasettet vårt innehar et tilfeldig utvalg av n observasjoner, som er gitt ved:

$$\{(x_{i1}, x_{i2}, \dots, x_{ik}, y_i) : i = 1, 2, \dots, n\} \quad (3.2)$$

Forutsetningen for tilfeldig utvalg sier at enhver observasjon i har den samme sannsynligheten for å bli valgt. I datasettet vårt har vi redusert antallet observasjoner for å utelukke resultater som ikke hadde den nødvendige informasjon for å kunne inkluderes. Datasettet vårt inkluderer likevel et tilfredsstillende antall observasjoner, og vi antar at MLR.2 holder.

MLR.3 Multikollinearitet

For at analysen skal holde forutsettes det også at antakelsen om ingen observert perfekt multikollinearitet er valid. Multikollinearitet er et mål på graden av lineær sammenheng mellom uavhengige variabler i en regresjonsmodell. Forutsetningen åpner for og antar korrelasjon mellom de uavhengige variablene i regresjonsmodellen, men hvis perfekt korrelasjon oppstår vil vi ikke lenger kunne estimere OLS med utgangspunktet i datasettet (Wooldridge, 2019, s.81).

Under Gauss-Markov forutsetningene, betinget av verdier fra et tilfeldig utvalg av de uavhengige variablene, har vi at variansen for helningen av OLS estimatorer er gitt ved:

$$\text{Var}(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1-R_j^2)} \quad (3.3)$$

Størrelsen på variansen til $\hat{\beta}_j$ sier noe om usikkerheten til estimatene våre, hvor en høy varians fører til bredere konfidensintervall og mindre presise estimater. Fra ligning (3.3) ser vi at variansen til OLS-estimatene blir påvirket av σ^2 , SST_j og R_j^2 hvor j kan betegne enhver av de uavhengige variablene (Wooldridge, 2019, s.89).

- Feilleddets varians, σ^2 :
Hvis variansen til feilleddet (σ^2) øker, vil variansen til OLS-estimatene også øke. Fordi σ^2 er et kjennetegn fra populasjonen og ikke datasettet, vil den være ukjent. For å redusere σ^2 vil man måtte legge til flere relevante faktorer fra feilleddet som uavhengige variabler i ligningen. Dette kan derimot være vanskelig å finne.
- Den totale variansen, SST_j :
Høy total variasjon i x_j vil gi en lav estimert varians for ($\hat{\beta}_j$). Gitt at alt annet holdes likt, ønsker vi derfor så mye variasjon i x_j som mulig. Ved et høyere antall observasjoner i datasettet vil variasjonen i hver av de uavhengige variablene øke systematisk deretter, og vi får dermed en lavere varians for OLS-estimatene.
- Forklaringsgrad mellom uavhengige variabler, R_j^2 :
 R_j^2 vil ligge en plass mellom 0 og 1, hvor en R_j^2 som går mot 1 indikerer at en uavhengig variabel forklarer mye av variasjonen i en annen uavhengig variabel, noe som vil si at de er sterkt korrelert. Høyere nivåer av R_j^2 gir høyere varians for OLS-

estimatene, men er ikke et brudd på forutsetningen. R_j^2 lik 1 gir perfekt korrelasjon, noe som vi har sett gjør at MLR.3 ikke holder.

Korrelasjonsmatrisen for regresjonsmodellen kan brukes til å sjekke for multikollinearitet i analysen.. For å gjøre dette settes en grenseverdi på for eksempel 90% som referansepunkt. Denne grenseverdien betyr ikke nødvendigvis at MLR.3 ikke holder, men det viser at det er variabler i modellen som har en sterk lineær sammenheng med andre variabler (Wooldridge, 2019, s.90).

For å unngå tilfellet med perfekt multikollinearitet kan man vurdere å droppe variabler fra modellen, men dette kan igjen føre til problemet med utelatte variabler hvis man ikke tar med faktorer som er nødvendige for å predikere modellen.

MLR.4 Forutsetning for eksogenitet

Forutsetningen om eksogenitet innebærer at restleddet u ikke skal korrelere med noen av de uavhengige variablene i modellen. Dette er nøkkelantakelsen for den generelle multiple regresjonsmodellen. Formelt er dette gitt ved:

$$E(u|x_1, x_2, \dots, x_{24}) = 0 \quad (3.4)$$

Formel 3.4 sier at det uobserverte restleddet, også kalt feilleddet, ikke er korrelert med de uavhengige variablene. Antakelsen viser også at vi har gjort riktig rede for forholdet mellom de uavhengige og avhengige variablene. Hvis u er korrelert med noen av de uavhengige variablene vil ligning (3.4) ikke holde. Med en multipl regressjonsanalyse kan vi inkludere mange uavhengige faktorer, og det er mindre sannsynlig at viktige faktorer utelates enn i en enkel regresjonsanalyse. Likevel vil det i alle modeller være faktorer som vi ikke kan ta hensyn til på grunn av begrensninger i datasettet eller ignoranse, og vil dermed utelukkes fra analysen. Hvis dette er faktorer som burde bli hensyntatt i analysen, vil MLR.4 ikke holde (Wooldridge, 2019, s.82).

I denne oppgaven antas det at MLR.1 til MLR.4 holder, og vi kan da si at den forklarende variabelen y_1 er eksogen.

MLR.5 Homoskedastisitet

Antakelsen om homoskedastisitet antar et konstant feilledd, og vi kan se hvorvidt OLS er effektivisert. I denne forutsetningen antar vi altså at feilleddet u vil ha samme varians for enhver verdi av de uavhengige variablene (Wooldridge, 2019, s.88). Dette kalles homoskedastisitet, og er gitt ved:

$$\text{Var}(u|x_1, \dots, x_k) = \sigma^2 \quad (3.5)$$

Hvis forutsetningen ikke holder, vil modellen ha heteroskedastisitet. Dvs. feilleddet vil avvike på tvers av de uavhengige variablene. For at forutsetningen skal holde kreves det at variansen til feilleddet ikke er avhengig av nivået på de uavhengige variablene x_1, x_2, \dots, x_k . Hvordan vi tester for dette går vi nærmere innpå i kapittel 3.3.2.

MLR.6 Normalfordeling av feilleddet

Den siste forutsetningen handler om normalfordeling av feilleddet u . Denne antakelsen sier at feilleddet u er uavhengig fra de forklarende variablene x_1, x_2, \dots, x_k og vil nærme seg en standard normal distribusjon med gjennomsnitt lik null og varians gitt ved:

$$u \sim N(0, \sigma^2) \quad (3.6)$$

MLR.6 er den strengeste antakelsen i CLM. Forutsetningen innebærer at MLR.4 og MLR.5 holder (Wooldridge, 2019, s.118). Det er derfor hensiktsmessig å bruke sentralgrenseteoremet for at modellen skal holde. Sentralgrenseteoremet sier at for et tilfeldig utvalg av en populasjon vil distribusjonen være tilnærmet normalfordelt. Jo høyere antall observasjoner, jo mer vil den ligne en standard normalfordelt distribusjon ifølge dette teoremet (Wooldridge, 2019, s.724).

I analysen er det gjort rede for en rekke faktorer som kan påvirke sannsynligheten for å fullføre studier på normert tid. Det er likevel flere faktorer fra datasettet som vi ikke har inkludert i analysen som kan ha en effekt på den avhengige variabelen. I analysen er det tatt med de mest relevante variablene, basert på tidligere forskning.

Vi antar at MLR.6 holder utfra sentralgrenseteoremet fordi datasettet har relativt mange observasjoner. En slik normalfordelt distribusjon i feilleddet u er også sentralt for å kunne utføre hypotesetester, noe vi vil analysere videre i kapittel 3.3.

3.2.5 Lineær sannsynlighetsmodell

I vår oppgave er den avhengige variabelen, y_1 , binær. Den multiple regresjonsmodellen vil dermed være en lineær sannsynlighetsmodell («linear probability model» forkortet til LPM).

$$P(y = 1|x) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k \quad (3.7)$$

Ligning 3.7 forklarer at sannsynligheten for suksess er $p(x) = P(y = 1|x)$, som er en lineær funksjon av x_j . Ligningen er en sannsynlighetsrespons. I en LPM modell vil β_j måle endringen i sannsynlighet for suksess når x_j endres, alt annet likt. Den estimerte ligningen blir da som følger:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_k x_k \quad (3.8)$$

Hvor \hat{y} er den estimerte sannsynligheten for suksess og $\hat{\beta}_0$ er den predikerte sannsynligheten for suksess når hver enkelt x_j er lik null. $\hat{\beta}_1$ måler den predikerte forandringen i sannsynligheten for suksess når x_1 øker med en enhet (Wooldridge, 2019, s.240).

3.3 Hypotesetesting

I denne undersøkelsen vil det testes hypoteser om parameterne i modellen. Ved å bruke hypotesetester vil det kunne undersøkes om det er dekkende statistisk bevis for om en påstand er sann, eller om det er nødvendig å forkaste påstanden til fordel for en alternativ påstand. For å kunne utføre en hypotesetest fremstilles det en nullhypotese, H_0 , og en alternativhypotese, H_1 , som er motsigende. Deretter brukes det en testobservator og en kritisk verdi som skal være med å avgjøre utfallet av testen. Videre oppgis det et signifikantnivå, α . Den skal gi oss en sannsynlighet for å kunne forkaste en gyldig hypotese. Det benyttes t-tester som skal teste hypotese på enkeltparametere og F-tester som skal teste hypotese på flere parametere samtidig. Testene baserer seg på antakelsene i kapittel 3.2.4 om at u er tilnærmet normalfordelt.

For å kunne teste effekten av enkeltparametere brukes t-testen. Her kan det for eksempel testes om $H_0: \beta_i = 0$ mot en alternativhypotese, $H_1: \beta_i \neq 0$. I en t-test må det tas i bruk en testobservator (TS) som er Student's t fordelt med $n-k-1$ frihetsgrader. Hvor n er antall observasjoner og k er parameter i modellen.

$$TS = \frac{\hat{\beta}_i - \beta_i}{\sigma} \sim t_{\frac{\alpha}{2}, n-k-1} \quad (3.9)$$

Her er $\hat{\beta}_i$ den estimerte verdien til β_i . β_i er verdien vi bruker under nullhypotesen og σ er det estimerte standardavviket assosiert med $\hat{\beta}_i$.

Ved en tosidig hypotesetest blir forkastningsregelen som følger: Vi forkaster H_0 dersom $|TS| > t_{\frac{\alpha}{2}, n-k-1}$. Det kan utføres en ensidig test på samme måte, men da skal ikke signifikantnivået deles på to.

Når det skal testes for felles hypotese om flere parametere samtidig må det brukes F-test. Dette kan for eksempel være en hypotese om at både β_1 og β_2 er lik 0 testes. Testen baserer seg på endringen i summen av kvadratisk residual gitt restriksjoner. Dersom det observeres en betydelig endring i SSR, er restriksjonene under nullhypotesen ugyldig.

$$TS = \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)} \sim F_{\alpha}(q, n-k-1) \quad (3.10)$$

Her er

SSR_r - summen av kvadratisk residual med restriksjoner

SSR_{ur} - summen av kvadratisk residual uten restriksjoner

q - antall restriksjoner under nullhypotesen

k - antall parameter I modellen

n - antall observasjoner

Kan også benytte determinasjonskoeffisienten, R^2 . Da skrives ligning 3.10 om.

$$TS = \frac{(R_{ur}^2 - R_r^2)/q}{(1 - R_{ur}^2)/(n-k-1)} \sim F_{\alpha}(q, n-k-1) \quad (3.11)$$

Her er

R_r^2 - determinasjonskoeffisienten med restriksjoner

R_{ur}^2 - determinasjonskoeffisienten uten restriksjoner

q - antall restriksjoner under nullhypotesen

k - antall parameter I modellen

n - antall observasjoner

Beslutningsregelen blir da å forkaste nullhypotesen, H_0 , dersom $|TS| > F_\alpha(q, n - k - 1)$.

3.3.2 Breusch-Pagan test

Ved å bruke Breusch-Pagan test vil det være mulig å kunne observere om det er heteroskedastisitet til stede i modellen. Dette gjøres ved å lage en nullhypotese om at det er konstant feilledd i modellen, det vil si homoskedastisitet. Alternativhypotesen er at det ikke er konstant feilledd, heteroskedastisitet. Testen ble gjennomført i Stata. Testen gir en output hvor den gir en verdi for testobservatoren og en p-verdi. P-verdien forteller for hvilket signifikansnivå vi kan forkaste nullhypotesen. Den vil ligge et sted mellom 0 og 1 hvor for eksempel 0.005 betyr et signifikantnivå på 5%. Dersom p-verdien er mindre enn signifikantnivået vil nullhypotesen forkastes og konkludere med det er heteroskedastisitet i modellen. Dersom det skjer, kan en bruke robuste standardavvik (Wooldridge, 2019, s.269-270).

3.4 Oppsummering

I dette kapitlet er det blitt redegjort for de ulike metodene som har blitt valgt for å gjennomføre denne undersøkelsen. Teoriene bak metodene er forklart og bakgrunnen til hvorfor de er brukt i denne oppgaven.

4 Presentasjon av datamateriale

4.1 Innledning

I dette kapitlet blir det redegjort for datamaterialet som er brukt. Først presenteres datamaterialet og undersøkelsen den er hentet fra. Videre blir det en gjennomgang av deskriptiv statistikk for avhengig variabel, interessevariabel og kontrollvariabler.

4.2 Datakilder

Oppgaven baseres på *Den europeiske studentundersøkelsen - Eurostudent VI* fra 2016, en nasjonal spørreundersøkelse. Datasettet er innhentet fra Norsk senter for forskningsdata AS (NSD), og undersøkelsen er gjennomført av Statistisk Sentralbyrå. Formålet med *Eurostudent VI* er å kartlegge sosiale og økonomiske vilkår for studenter i høyere utdanning, med fokus på studentenes sosiale bakgrunn, levekår og tidsbruk. (Statistics Norway, 2022).

4.3 Deskriptiv statistikk for datamaterialet

I denne regresjonsanalysen er den avhengige variabelen y_1 en dummyvariabel som uttrykker om studenter har planer om å fortsette utdanningen sin eller ikke. Vi har en interessevariabel, *Arbeidstid* som angir antall timer respondentene jobber per uke. Videre brukes følgende kontrollvariabler : *alder_u22*, *alder22_24*, *alder25_29*, *female*, *mig1*, *mig2*, *morsmål*, *velstand*, *middelsvelstand*, *pared*, *barn*, *bachelor*, *master*, *profesjon*, *human*, *samf*, *helse*, *it*, *naturvit*, *lærer*, *ingeniør*, *bedøk*

Datamaterialet vi benytter oss av omfatter kun norske studenter. Tabell 1 gir en oversikt over alle variablene vi har benyttet i vår analyse.

Beskrivelse av variabler

Variabelnavn	Forklaring	Variabeltype
<i>y1</i>	Studenten skal fortsette på pågående studie = 1, 0 ellers	Dummy
<i>Arbeidstid</i>	Antall timer med lønnet arbeid i uken	Kontinuerlig
<i>alder_u22, alder22_24, alder25_29</i>	Aldersgrupper = 1, 0 ellers	Dummy
<i>female</i>	Kvinne=1, mann= 0	Dummy
<i>barn</i>	Studenten har barn=1, 0 ellers	Dummy
<i>mig1</i>	Studenten er førstegenerasjonsmigrant=1, 0 ellers	Dummy
<i>mig2</i>	Studenten er andregenerasjonsmigrant = 1, 0 ellers	Dummy
<i>velstand, middelsvelstand</i>	Velstandsnivå studenten kommer fra = 1, 0 ellers	Dummy
<i>morsmål</i>	Norsk som morsmål = 1, 0 ellers	Dummy
<i>bachelor, master, profesjon</i>	utdanningsnivå =1, 0 ellers	Dummy
<i>lærer, human, samf, bedøk, naturvit, it, ingeniør, helse</i>	Fagfelt studenten tilhører = 1, 0 ellers	Dummy

Tabell 1: oversikt over alle variabler

4.3.1 Deskriptiv statistikk, avhengig variabel

Den avhengige variabelen, y_1 , en dummyvariabel som i utgangspunktet er et spørsmål i undersøkelsen studentene svarte på. Spørsmålet lyder som følger: «Planlegger du å fortsette studiene etter å ha fullført programmet/ene du nå følger?». De hadde fire alternativer de kunne svare: 1) «Ja, innen ett år», 2) «Ja, men ikke innen ett år», 3) «Nei, jeg har ingen planer om videre studier» og 4) «vet ikke». Svarkategori 1) og 2) er slått sammen og gitt verdien = 1.

Svarkategoriene 3) og 4) er gitt verdien = 0. *Tabell 2* nedenfor gir en oversikt over fordelingen i de to kategoriene.

Deskriptiv statistikk, avhengig variabel

<i>y₁</i>		<i>Frekvens</i>	<i>Prosent</i>
Antall observasjoner	8 235		
Avslutte	0	3 988	49.02%
Fortsette	1	4 148	50.98%

Tabell 2: Deskriptiv statistikk for avhengig variabel, y_1 hvor «Fortsette» = studenter som planlegger å fortsette studiet, «Avslutte» = studenter som planlegger å avslutte studiet

I datasettet var det opprinnelig 8 235 observasjoner. Dersom variabelen har en verdi på -99 utelukkes den, det samme gjelder for alle resterende variabler i analysen. . I tillegg blir alle variabler med verdi over 50 timer utelukket da vi ser det som usannsynlig at noen jobber så mye ved siden av studiene. Etter å ha rensket datasettet ble det igjen 8 136 observasjoner igjen. I datasettet er det en relativt jevn fordeling mellom de som planlegger å fortsette og avslutte studiene. Likevel er det en litt høyere andel som tenker å fortsette.

4.3.2 Deskriptiv statistikk for interessevariabel

Interessevariabelen er *arbeidstid* og det er en kontinuerlig variabel. Her opplyser studentene antall timer lønnet arbeid per uke.. *Tabell 3* viser at vi har 8 136 observasjoner med et gjennomsnitt på 11 lønnet arbeidstimer. Standardavvik ligger på 12 som er et relativt høyt mål, som betyr at det er høy spredning mellom antall arbeidstimer studentene jobber. Vi har en median lik 8. Her ser vi at gjennomsnittet og medianen avviker fra hverandre, noe som indikerer at gjennomsnittet ikke er representativt for midten i fordelingen.

Deskriptiv statistikk, interessevariabel

<i>Arbeidstid</i>	
Antall observasjoner	8 136
Gjennomsnitt	11.34
Standardavvik	12.91
Median	8
Min	0
Maks	50

Tabell 3: Deskriptiv statistikk for interessevariabel Arbeidstid, den viser arbeidstimer per uke

Histogram av interessevariabel Arbeidstid

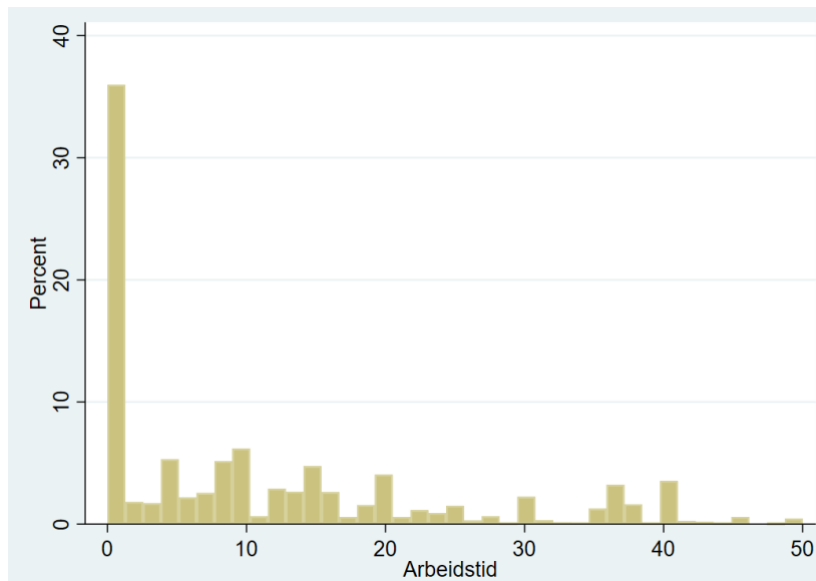


Diagram 1: Prosentvis fordeling av interessevariabel, Arbeidstid

Ser fra *Diagram 1* at en betydelig del av studentene, 35.32%, oppgir arbeidstid=0. Dette betyr at de ikke har lønnet arbeid i forelesningsperioden.

4.3.3 Deskriptiv statistikk av kontrollvariabler

Kontrollvariablene vi har valgt å bruke går under sosiodemografiske kjennetegn til individet og kjennetegn ved utdanningen de tar. Som sosiodemografiske kjennetegn har vi valgt å bruke individets aldersgruppe, kjønn, morsmål, velstandsnivå, migrasjon, foreldres utdanningsnivå og om de har barn. Som utdanningsvariabler har vi valgt å dekke utdanningsnivå og fagområdet. Kontrollvariabler som skal dekke individets utdanningsnivå skiller mellom bachelor-, 2-årig master og profesjonsstudium som også inkluderer 5-årig integrert master. For fagområde har vi valgt å skille mellom de som går følgende: lærerutdanning/pedagogikk, humanetiske/estetiske fag, samfunnsfag/journalisme/biblioteks- og arkivs fag, bedriftsøkonomiske/administrative/juridiske fag, naturvitenskapelige/matematiske/statistiske fag, informasjons- og datateknologi, tekniske-/ingeniør-/håndverksfag, landbruk/fiskeri/skogsbruk/veterinær, helse- og sosialfag, transport-/sikkerhetsfag og personlig tjenesteyting.

alder_u22, *alder22_24*, *alder25_29* er dummyvariabler som tar verdien 1 dersom individet enten er under 22, mellom 22 og 24 eller mellom 25 og 29. Variabelen tar verdien 0 hvis ikke. Referanse kategorien er de over 30 år.

Deskriptiv statistikk, aldersgruppe

Variabel	Frekvens	Prosent
<i>alder_u22</i>	1 724	21%
<i>alder22_24</i>	2 552	31%
<i>alder25_29</i>	1 940	24%
<i>alder_over30 (ref.kategori)</i>	1 920	24%
Totalt	8 136	100%

Tabell 4: Deskriptiv statistikk for kontrollvariablene, *alder_u22*, *alder22_24*, *alder25_29*

Female er en dummyvariabel hvor den tar verdien 1 dersom individet er kvinne og verdien 0 dersom individet er mann.

Deskriptiv statistikk, kjønn

<i>female</i>	Frekvens	Prosent
1	5 044	62%
0	3 092	38%
Totalt	8 136	100%

Tabell 5: Deskriptiv statistikk for kontrollvariabel, *female*

Morsmål er en dummyvariabel som beskriver om individet har norsk som morsmål eller ikke. Den får verdien 1 dersom individet norsk som morsmål og 0 hvis ikke.

Deskriptiv statistikk, morsmål

<i>morsmål</i>	Frekvens	Prosent
1	7 404	91%
0	732	9%
Totalt	8 136	100%

Tabell 6: Deskriptiv statistikk for kontrollvariabel, *morsmål*

Ved individets velstandsnivå har vi valgt å skille mellom høy velstand, middelsvelstand og lav velstand. Det betyr at verdiene med høy velstand og velstand blir slått sammen og får dummyvariabelen *velstand*. Den tar verdien 1 dersom de anser foreldrene sin økonomiske situasjon som velstående eller høy velstand og 0 hvis ikke. *middelsvelstand* tar verdien 1 dersom individet anser foreldrene sine økonomiske situasjon som gjennomsnittlig og verdien 0 hvis ikke. Referansekategori er lav velstand og svært lav velstand.

Deskriptiv statistikk, velstandsnivå

Variabel	Frekvens	Prosent
<i>Velstand</i>	2 798	34%
<i>Middelsvelstand</i>	4 512	56%
<i>lav velstand (ref.kategori)</i>	826	10%
Totalt	8 136	100%

Tabell 7: Deskriptiv statistikk for kontrollvariablene, velstand, middelsvelstand

Ved individets migrasjonsbakgrunn skiller vi mellom første- og andregenerasjonsinnvandrere.

Vi laget dummyvariabelen *mig1* som tar verdien 1 dersom individet er første generasjonsinnvandrer og 0 hvis ikke. *mig2* er en dummyvariabel som tar verdien 1 dersom de er andregenerasjonsinnvandrer og 0 hvis ikke. Referansekategorien er dersom individet er født i Norge med to norskfødte foreldre.

Deskriptiv statistikk, innvandringsbakgrunn

Variabel	Frekvens	Prosent
<i>mig1</i>	811	10%
<i>mig2</i>	740	9%
<i>norskfødte foreldre (ref.kategori)</i>	6 585	81%
Totalt	8 136	100%

Tabell 8: Deskriptiv statistikk for kontrollvariablene, mig1, mig2

pared er en dummyvariabel som beskriver foreldrenes utdanningsnivå. Den tar verdien 1 dersom foreldrene har utdanning på universitets- eller høgskolenivå og tar verdien 0 hvis ikke.

Deskriptiv statistikk, foreldres utdanningsnivå

<i>pared</i>	Frekvens	Prosent
1	5 867	28%
0	2 269	72%
Totalt	8 136	100%

Tabell 9: Deskriptiv statistikk for kontrollvariabelen, *pared*

barn er en dummyvariabel som beskriver om individet har barn eller ikke. Den tar verdien 1 dersom de har barn og verdien 0 hvis ikke.

Deskriptiv statistikk, barn

<i>barn</i>	Frekvens	Prosent
1	1 588	20%
0	6 548	80%
Totalt	8 136	100%

Tabell 10: Deskriptiv statistikk for kontrollvariabelen, barn

Ved individets utdanningsnivå skiller vi mellom bachelor, master og profesjonsstudium, og har laget en dummyvariabel for hver av disse. Dummyvariabelen *bachelor* tar verdien 1 dersom de går på et bachelorstudium og verdien 0 hvis ikke. Dummyvariabelen *master* beskriver om studenten går på et 2-årig masterstudium og tar verdien 1 dersom de går master og verdien 0 hvis ikke. Dummyvariabelen *profesjon* beskriver om studenten går på en 5-årig integrert master eller 6-årig profesjonsstudium og tar verdien 1 dersom de går profesjon og 0 hvis ikke. Referansekategori fanger opp om studenten går årsstudium eller noe annet som inkluderer enkeltemner, den viser da verdien 0.

Deskriptiv statistikk, utdanningsnivå

Variabel	Frekvens	Prosent
<i>bachelor</i>	4 204	52%
<i>master</i>	2 128	26%
<i>profesjon</i>	553	7%
<i>årsstudium (ref.kategori)</i>	1 251	15%
Totalt	8 136	100%

Tabell 11: Deskriptiv statistikk for kontrollvariablene, bachelor, master, profesjon

Ved individets fagområde har vi gjort de ulike fagfeltene til dummyvariabler. Lærerutdanning/pedagogikk er gjort om til dummyvariabelen *lærer*, som viser verdien 1 dersom individets utdanning går under fagfeltet lærerutdanning/pedagogikk og gir verdien 0 hvis ikke. Dummyvariabelen *human* beskriver om utdanningen går under fagfeltet humanetiske/estetiske fag og gir verdien 1 dersom de gjør det og 0 hvis ikke.

Dummyvariabelen *samf* beskriver om utdanningen går under fagfeltet samfunnsfag/journalisme/biblioteks- og arkivfag og gir verdien 1 dersom de gjør det og 0 hvis ikke. Dummyvariabelen *bedøk* beskriver om utdanningen går under fagfeltet bedriftsøkonomiske/administrative/juridiske fag og gir verdien 1 dersom de gjør det og 0 hvis ikke. Dummyvariabelen *naturvit* beskriver om utdanningen går under fagfeltet naturvitenskapelige/matematiske/statistiske fag og gir verdien 1 dersom de gjør det og verdien 0 hvis ikke. Dummyvariabelen *it* beskriver om utdanningen går under fagfeltet informasjons- og datateknologi og gir verdien 1 dersom de gjør det og verdien 0 hvis ikke. Dummyvariabelen *ingeniør* beskriver om utdanningen går under fagfeltet tekniske-/ingeniør-/håndverksfag og gir verdien 1 dersom de gjør det og verdien 0 hvis ikke. Dummyvariabelen *helse* beskriver om utdanningen går under fagfeltet helse- og sosialfag, og viser verdien 1 dersom de gjør det og verdien 0 hvis ikke. Referansekategorien er fagfeltet landbruk/fiskeri/skogsbruk/veterinær og fagfeltet transport-/sikkerhetsfag og personlig tjenesteyting.

Deskriptiv statistikk, fagområde

Variabel	Frekvens	Prosent
<i>lærer</i>	1283	16%
<i>human</i>	681	8%
<i>samf</i>	744	9%
<i>bedøk</i>	1 561	19%
<i>naturvit</i>	596	7%
<i>it</i>	352	4%
<i>ingeniør</i>	982	12%
<i>helse</i>	1 534	19%
<i>landbruk_transport (ref.kategori)</i>	403	5%
Totalt	8 136	100%

Tabell 12: Deskriptiv statistikk for kontrollvariablene som beskriver fagfelt

4.5 Oppsummering

Dette kapittelet redegjør for datamaterialet. Det beskriver deskriptiv statistikk for alle variablene som skal være med i analysen. Til tross for en stor mengde kontrollvariabler vil det alltid eksistere uobserverbare variabler som kan være korrelert med de inkluderte og som påvirker utdanningsutfallet.

5 Empiriske resultater

5.1 Innledning

I denne delen skal vi analysere sammenhengen mellom utdanningsutfall og lønnet arbeidstid ved bruk av datamaterialet som er blitt presentert. I analysen brukes minste kvadraters metode for en lineær sannsynlighetsmodell og hypotesetesting. Først skal vi ta for oss en enkel lineær modell med valgt avhengig variabel og interessevariabelen *Arbeidstid*. Deretter utvides modellen med kontrollvariabler. Inkludering av kontrollvariabler gjøres i to steg: 1) Inklusjon av kontrollvariabler som skal kontrollere for individets bakgrunn og 2) inklusjon av variabler som uttrykker individets utdanningsbakgrunn. Videre utvides modellen for å se om det er noen ikke-lineære sammenhenger. Regresjonsanalysen er gjort med OLS-metoden i dataprogrammet Stata. For å sørge for at vi har restledd som varierer systematisk vil vi teste modellen vår for heteroskedastisitet. Til slutt skal vi diskutere våre resultater opp mot tidligere funn.

5.2 Resultater

Dette kapittelet presenter resultatene av regresjonsanalysene. Modell i) er en enkel lineær regresjon, modell ii) utvidet lineær modell med kontroll for sosiodemografiske variabler, modell iii) utvidet lineær modell med kontroll for utdanningsvariabler og modell iv) utvidet ikke-lineær modell.

5.2.1 Modell i: enkel lineær modell

Den første modellen er en enkel lineær modell. I denne modellen kan vi se hvordan studentenes planer om å fullføre studiene sine blir påvirket av antall lønnet arbeidstimer de har i forelesningsperioden. Modellen er gitt ved:

$$y_1 = \beta_0 + \beta_1 \text{Arbeidstid} + u \quad (i)$$

Ved bruk av Stata utfører vi en regresjonsanalyse med minste kvadraters metode. Tabell 13 andre kolonne viser estimatorene til koeffisienten i modell *i*. Tallene i parentes viser de estimerte standardavvikene til estimatorene.

Regresjonsmodeller

<i>Modell</i>	i	ii	iii	iv
VARIABLES	y1	y1	y1	y1
<i>Arbeidstid</i>	-0.00205*** (0.00043)	0.00100** (0.00049)	0.00049 (0.00048)	0.00090 (0.00128)
<i>Arbeidstid_sqrd</i>				-0.00001 (0.00003)
<i>alder_u22</i>		0.30049*** (0.02253)	0.21731*** (0.02186)	0.21690*** (0.02189)
<i>alder22_24</i>		0.20792*** (0.02090)	0.16948*** (0.02023)	0.16876*** (0.02034)
<i>alder25_29</i>		0.07736*** (0.02002)	0.08140*** (0.01914)	0.08062*** (0.01927)
<i>mig1</i>		0.03929 (0.03164)	0.04028 (0.03002)	0.04016 (0.03002)
<i>mig2</i>		0.02127 (0.01932)	0.02208 (0.01834)	0.02204 (0.01834)
<i>female</i>		-0.00202 (0.01135)	-0.03787*** (0.01135)	-0.03810*** (0.01137)
<i>velstand</i>		-0.03602* (0.02049)	-0.01593 (0.01947)	-0.01594 (0.01947)
<i>middelsvelstand</i>		-0.01929 (0.01910)	-0.01657 (0.01812)	-0.01651 (0.01813)
<i>pared</i>		0.00224 (0.01269)	0.02408** (0.01212)	0.02404** (0.01212)
<i>barn</i>		0.05190*** (0.01957)	0.03782** (0.01867)	0.03821** (0.01870)
<i>morsmål</i>		-0.04374 (0.03292)	-0.05690* (0.03124)	-0.05685* (0.03124)
<i>bachelor</i>			0.02558 (0.01665)	0.02483 (0.01679)
<i>master</i>			-0.31457*** (0.01757)	-0.31481*** (0.01759)
<i>profesjon</i>			-0.18904*** (0.02562)	-0.18966*** (0.02569)
<i>human</i>			0.02799 (0.02956)	0.02757 (0.02958)
<i>samf</i>			0.00660 (0.02915)	0.00616 (0.02918)
<i>helse</i>			-0.02316 (0.02665)	-0.02371 (0.02670)
<i>it</i>			-0.15073*** (0.03469)	-0.15076*** (0.03469)
<i>naturvit</i>			0.04583 (0.03049)	0.04568 (0.03049)
<i>lærer</i>			-0.04731* (0.02697)	-0.04769* (0.02700)
<i>ingenior</i>			-0.15030*** (0.02829)	-0.15028*** (0.02829)
<i>bedøk</i>			-0.05269** (0.02646)	-0.05302** (0.02648)
<i>Constant</i>	0.53312*** (0.00737)	0.39767*** (0.04130)	0.56700*** (0.04664)	0.56716*** (0.04665)
<i>Observations</i>	8,136	8,136	8,136	8,136
<i>R-squared</i>	0.003	0.036	0.135	0.135

Tabell 13: Regresjonsmodell i, ii, iii og iv. Standardavvik i parentes. ***p<0.01, ** p>0.05, *p<0.1

Homoskedastisitettesten, Breusch-Pagan, viser at testobservatoren er på 0.24 med tilhørende p-verdi 0.6276. Basert på dette er det ikke nok bevis til å kunne forkaste nullhypotesen om homoskedastiske restledd i modell i for et signifikansnivå lik 1%.

Fra Tabell 13 kolonne 2, observeres det at konstantleddet for modell i er 0.53312. Dette indikerer at dersom studenten ikke jobber, vil sannsynligheten for at de planlegger å fortsette studiene være lik 0.53312. Dersom studenten har jobb, vil antall arbeidstimer påvirke utfallsvariabelen negativt. En økning i antall arbeidstimer vil minske sannsynligheten for å fortsette studiene med 0.00205. Når modellen utvides, skal denne effekten studeres nærmere. R^2 er målt til 0.004, som betyr at denne modellen forklarer 0.4% av variasjonen i $y1$. Variabelen *Arbeidstid*, er signifikant ($p < 0,01$), men negativ koeffisient. Dette innebærer at en ekstra lønnet arbeidstime i uken vil redusere sannsynligheten for at studenten planlegger å fortsette med studiene med 0.00205, alt annet likt.

5.2.2 Modell ii: utvidet lineær modell med kontroll for sosiodemografiske variabler

Variablene i modell ii er kontrollvariabler som skal minske sannsynligheten for utelatt variabelskjevhet. Dette kommer som følge av at interessevariabelen kan fange opp variasjonene av andre variabler som ikke er tatt med i modellen. Kontrollvariabler vil sørge for mer korrekte estimatorer for effekten av interessevariabelen. Kontrollvariablene skal dekke individets sosiodemografiske bakgrunn. Kontrollvariablene som er valgt å ha med i modellen er følgende: *female*, *alder_u22*, *alder22_24*, *alder25_29*, *mig1*, *mig2*, *morsmål*, *velstand*, *middelsvelstand*, *pared* og *barn*.

Med disse kontrollvariablene får vi følgende modell ii:

$$y1 = \beta_0 + \beta_1 \text{Arbeidstid} + \beta_2 \text{alder_u22} + \beta_3 \text{alder22_24} + \beta_4 \text{alder25_29} + \beta_5 \text{mig1} + \beta_6 \text{mig2} + \beta_7 \text{female} + \beta_8 \text{velstand} + \beta_9 \text{middelsvelstand} + \beta_{10} \text{pared} + \beta_{11} \text{barn} + \beta_{12} \text{morsmål} \quad (\text{ii})$$

Breusch-Pagantesten viser en testobservatør på 0.44 med tilhørende p-verdi 0.5094. Dette betyr at det ikke er nok bevis til å kunne forkaste nullhypotesen om konstant restleddvarians med signifikansnivå på 1%. Kan ut ifra dette anta at det er homoskedastisitet i modell ii.

Resultatene for modell ii er gitt i Tabell 13 kolonne 3. Ved å inkludere de sosiodemografiske kontrollvariablene endres effekten av interessevariabelen *Arbeidstid* noe. Også i denne modellen ble *Arbeidstid* signifikant ($p < 0,001$) og koeffisienten = 0.001. Dette indikerer at ved en økning i arbeidstid med en time, forventes det at sannsynligheten for at studenten fortsetter på studiet øker med 0.001, alt annet likt. Effekten av arbeidstiden har dermed økt til en positiv effekt etter at kontrollvariablene er lagt inn.

Fra kontrollvariablene for alder observeres det at individer under 22 år har høyest sannsynlighet for å fortsette studiene sine sammenlignet med de som er eldre enn 22 år gammel. Er studenten førstegenerasjonsinnvandrere er sannsynligheten 0.03929 høyere sammenlignet med om de ikke er det. Er de andregenerasjonsinnvandrere er koeffisienten noe lavere, 0.02127. Er studenten kvinne er sannsynlighet for at de fortsetter studiet 0.00202 lavere enn om de hadde vært mann. Har individet familiær bakgrunn med middelsvelstandsnivå vil sannsynligheten for at de fortsetter være 0.01929 lavere enn om de ikke kom fra en slik bakgrunn. Dersom foreldrene har utdanning på universitets- eller høyskolenivå er sannsynligheten 0.00224 høyere for at studenten planlegger å fortsette studiene sammenlignet med studentene som ikke har foreldre med høyere utdanning. Har studenten barn er sannsynligheten 0.05190 høyere for at de fortsetter sammenlignet med de som ikke har barn. Dersom studenten har norsk som morsmål er sannsynligheten 0.04374 lavere for at de fortsetter sammenlignet med om de ikke snakket norsk.

Determinasjonskoeffisienten, R^2 , har økt betraktelig til 3,6% etter at vi har lagt til kontrollvariablene. Det betyr at vi har fått en større forklaringskraft i modell ii.

Videre er *Arbeidstid* signifikant med 95% statistisk sannsynlighet. Variablene for alder og barn er signifikante med 99% statistisk sannsynlighet og *velstand* er signifikant med 90% statistisk sannsynlighet. Selv om de andre variablene viser seg å ikke være signifikante er det fortsatt relevant å ha dem med i modellen. De ikke-signifikante variablene vil fange opp noe av effekten interessevariabelen har på utfallsvariabelen (yI). På den måten reduseres risikoen for utelatt variabel skjevhet.

5.2.3 Modell iii: utvidet modell med kontroll for utdanningsvariabler

I denne modellen legges det til flere kontrollvariabler. Kontrollvariablene skal dekke individets utdanningsbakgrunn, og er som følger: *master*, *profesjon*, *human*, *samf*, *helse*, *it*, *naturvit*, *lærer*, *ingeniør*, *bedøk*, og får følgende modell:

$$y_1 = \beta_0 + \beta_1 \textit{Arbeidstid} + \beta_2 \textit{alder_u22} + \beta_3 \textit{alder22_24} + \beta_4 \textit{alder25_29} + \beta_5 \textit{mig1} + \beta_6 \textit{mig2} + \beta_7 \textit{female} + \beta_8 \textit{velstand} + \beta_9 \textit{middelsvelstand} + \beta_{10} \textit{pared} + \beta_{11} \textit{barn} + \beta_{12} \textit{morsmal} + \beta_{13} \textit{bachelor} + \beta_{14} \textit{master} + \beta_{15} \textit{profesjon} + \beta_{16} \textit{human} + \beta_{17} \textit{samf} + \beta_{18} \textit{helse} + \beta_{19} \textit{it} + \beta_{20} \textit{naturvit} + \beta_{21} \textit{lærer} + \beta_{22} \textit{ingeniør} + \beta_{23} \textit{bedøk} \quad (\text{iii})$$

Det testes også her for heteroskedastisitet ved en Breusch-Pagan test. Testobservatøren er her lik 4.99, med tilhørende p-verdi 0.0255. Det vil dermed ikke være nok bevis til å forkaste nullhypotesen med 1% signifikantnivå og konkluderer med at modell iii har homoskedastisitet.

For resultatene i modell iii henvises det til kolonne 4 i Tabell 13. Ved å inkludere flere kontrollvariabler vil effekten av de variablene som skal kontrollere for de sosiodemografiske kjennetegnene til individet ha minsket noe. Interessevariabelen *Arbeidstid* har fått større positiv effekt når det kontrolleres for flere variabler. Dersom individet tar en bachelorgrad vil sannsynligheten for at de fortsetter på studie være 0.02558 mer sammenlignet med om de gikk en annen grad. Går individet et toårig masterprogram er sannsynligheten 0.31457 mindre sammenlignet med de andre gradene. Dersom studenten går et profesjonsstudium på 5 eller 6 år er effekten 0.18904 mindre for at de fortsetter sammenlignet om de går en annen grad. For en student som tar en utdanning innenfor helse- og sosialfag, informasjons- og datateknologi, lærerutdanning/pedagogikk, tekniske- og ingeniørfag, og bedriftsøkonomiske- og administrative fag vil de hver for seg gi en lavere sannsynlighet for å fortsette studiet når man sammenligner de med andre fagfelt. Determinasjonskoeffisienten er i denne modellen på 0.135. Det vil si at forklaringskraften ligger på 13.5% når de resterende kontrollvariabler inkluderes.

Videre undersøkes det om interessevariabelen *Arbeidstid* er av betydning. Ifølge Tabell 13 observeres det at den ikke er signifikant. Kontrollvariablene for alder er fortsatt av betydning med 99% statistisk sannsynlighet. Kontrollvariabel *female* er nå blitt signifikant med 99%

statistisk sannsynlighet. *pared* og *barn* er av betydning med 95% statistisk sannsynlighet, mens *morsmål* er av betydning med 90% statistisk sannsynlighet. Utdanningsvariablene *master* og *profesjon* er signifikant med 99% statistisk sannsynlighet. Av de variablene som skal representere de ulike fagfeltene er det *it*, *lærer*, *ingeniør* og *bedøk* som er av betydning. I denne modellen er det også relevant å beholde alle variablene selv om de ikke er av betydning, som nevnt i kapittel 5.2.2.

5.2.4 Modell iv: utvidet ikke-lineær modell

Frem til nå har modellene hatt lineære effekter av arbeidstimer. Modell iii blir utvidet for å se om det er noen ikke-lineære effekter av arbeidstimer. Dette gjøres ved å legge til en variabel som er kvadrert i tillegg til den variabelen som er lineær. Denne variabelen får navnet *Arbeidstid_sqrd* og blir lagt til i modell iii, som gir modell iv.

$$y_1 = \beta_0 + \beta_1 \text{Arbeidstid} + \beta_2 \text{Arbeidstid_sqrd} + \beta_3 \text{alder_u22} + \beta_4 \text{alder22_24} + \beta_5 \text{alder25_29} + \beta_6 \text{mig1} + \beta_7 \text{mig2} + \beta_8 \text{female} + \beta_9 \text{velstand} + \beta_{10} \text{middelsvelstand} + \beta_{11} \text{pared} + \beta_{12} \text{barn} + \beta_{13} \text{morsmal} + \beta_{14} \text{bachelor} + \beta_{15} \text{master} + \beta_{16} \text{profesjon} + \beta_{17} \text{human} + \beta_{18} \text{samf} + \beta_{19} \text{helse} + \beta_{20} \text{it} + \beta_{21} \text{naturvit} + \beta_{22} \text{lærer} + \beta_{23} \text{ingeniær} + \beta_{24} \text{bedøk} \quad (\text{iv})$$

Ved Breusch-Pagan testen vil testobservatøren være lik 4.85, med tilhørende p-verdi 0.0276. Det kan konkluderes med at det er homoskedastisitet i modellen med 99% statistisk sannsynlighet.

Ved å sammenligne estimatene fra modell iv med de fra modell iii vil det kunne observeres at de fleste koeffisientene er tilnærmet like. Interessevariabelen *Arbeidstimer* (β_1) vil ha den største endringen. Koeffisienten til variabel *Arbeidstid_sqrd* er negativ, som betyr at effekten av arbeidstid er positiv, men i avtagende grad. Det betyr at modellen er konkav og har et toppunkt. Dette indikerer at sannsynligheten for at studenten fortsetter med studiene sine øker frem til det når dette toppunktet. Når arbeidstiden overstiger toppunktet, vil denne sannsynligheten minske. Toppunktet vil vise hvor mye arbeidstid ved siden av studiene som vil gi den største sannsynligheten for at studenten velger å fortsette studiene sine

Tabell 13 viser at modell iv og modell iii har de samme signifikante variablene, og dermed blir konklusjonene om signifikans den samme.

Vedlegg 1 viser korrelasjonsmatrisen for modell iv. Som nevnt i kapittel 3.2.4 er det satt en grenseverdi på 90% som referansepunkt, og fra matrisen vil det kunne observeres at det generelt er lave verdier av kollinearitet i analysen. Likevel er det et par variabler som har et sterkt lineært forhold med andre variabler, slik som *middelsvelstand* og *velstand*. Fra teorien om multikollinearitet kommer det fram at selv om det er variabler som ligger rundt referansepunktet, betyr ikke dette at MLR.3 ikke vil holde. For å unngå mulig perfekt multikollinearitet kunne det blitt vurdert å droppe noen av variablene som korrelerer sterkt med andre. I denne oppgaven er det valgt å ikke gjøre dette da disse variablene kan være nødvendige for å redusere muligheten for skjevhet i modellen. Det konkluderes med at det ikke vil være noen særlige problemer knyttet til multikollineariteten mellom disse variablene.

5.3 Diskusjon

Denne oppgaven analyserer hvilke faktorer som kan ha påvirkning for om en student velger å ikke fortsette studiet de har påbegynt, hvor arbeidstid er interessevariabel.

Modell i indikerer at arbeidstid vil ha en negativ effekt på utdanningsutfallet. Dette stemmer delvis med Hovdhaugen (2015) sin artikkel om frafall, hvor det konkluderes med at det å ha jobb ved siden av studie vil øke sannsynligheten for å droppe ut hvis studenten jobber over et bestemt antall timer.

I modell ii ble effekten av arbeidstid positiv. Den positive effekten av arbeidstid kan forklares med at de studentene som har ressurser til å arbeide under studietiden vil ikke slite økonomisk og dermed ha bedre forutsetninger til å fortsette studiene. Det er kjent fra aviser og offentlige diskusjoner at studenter hevder ordinært studielån ikke er nok til å dekke levekostnader (Katanasho, 2022). Studenter som starter et studieløp, er bevisst på at de må arbeide og har ressurser til å både håndtere å arbeide og samtidig studere. Dermed vil ikke arbeidstid påvirke frafall fra studie.

Resultatene fra regresjonen viser at eldre studenter slutter oftere enn yngre studenter, (modell ii). Dette kan forklares med at eldre studenter har i større grad mer økonomiske forpliktelser

enn det yngre studenter har. Noen har kjøpt bolig og kravet til levestandard og forbruksnivå kan være økende med økende alder. Det er rimelig å anta at eldre studenter er mer uavhengige av foreldre. Dette innebærer også at de ikke kan basere seg på økonomisk støtte fra foresatte. Dette medfører at denne gruppen med studenter vil måtte arbeide relativt mer for å opprettholde levestandarden. Det kan være at yngre studenter får mer støtte hjemmefra, som fører til at de kan være bedre rustet dersom noe uforventet skjer. Samtidig er det rimelig å anta at denne studentgruppen i mindre grad har økonomiske forpliktelser. Dette er sammenfallende med rapporterte funn i Hovdhaugen (2019) hvor forfatteren viser at studentenes sosiale bakgrunn har betydning for studiefrafallet.

Videre viser modell ii at dersom studenten har familiær bakgrunn med høyt velstandsnivå er det lavere sannsynlighet for at de fortsetter studiet. Dette kan forklares med at studenter med slik bakgrunn kan ha lavere terskel for å avslutte studie. Studenter fra hjem med høy velstand kan ha lavere behov for utdanning sammenlignet med andre studenter uten en slik familiær bakgrunn. I tillegg kan det argumenteres for at studenter med høy velstand ikke håndterer motstand og utfordringer en kan møte på i høyere utdanning sammenlignet med andre studenter. Bakgrunnen for dette kan være at de ikke har hatt økonomiske problemer og muligens tar mer ting for gitt.

Modell ii viser at dersom studenten har barn vil det gi en positiv effekt på utdanningsutfallet. Man kunne tenkt at barn vil gi en negativ effekt på utdanningsutfallet da barn krever en del økonomiske og tidkrevende ressurser. På en annen side kan man argumentere for den positive effekten i modell ii med at studenter med barn må i større grad være mer strukturert og målrettet for å kunne gjennomføre studie sammenlignet med de som ikke har barn.

Modell ii ble utvidet ved å legge til flere kontrollvariabler og kalt modell iii. Denne gangen ble det også kontrollert for utdanningsbakgrunnen til studenten. I denne modellen ble det observert at den positive effekten av arbeidstiden blir halvert. Dette kan være et resultat av at kontrollvariablene fanget opp mer av effekten på utdanningsutfallet.

Denne modellen har flere variabler signifikante. Modell iii viser at dersom studenten er kvinne er det en signifikant negativ effekt på utdanningsutfallet. Denne effekten er motsatt av Hovdhaugen sine funn, der de finner at menn i større grad har risiko for frafall enn kvinner. I

rapporten til Andresen og Lervåg (2022) ble det observert den samme effekten som i Hovdhaugen sin analyse. De motstridende effektene funnet i vår modell og deres analyser kan være et resultat av at forklaringsgraden, R^2 , er lav i modell iii.

Foreldres utdanningsnivå er det, ifølge modell iii, en indikasjon på en positiv effekt på utdanningsutfallet. Hovdhaugen (2015) kommer frem til samme resultat, at dersom begge foreldre har høy utdanning vil det minske risikoen for frafall. Dette er ikke utenkelig da ressurssterke foreldre i større grad kan støtte studentene økonomisk og til studiene. Den teoretiske betraktning er at for å maksimere nytten av utdanning, er foreldres påvirkning en av komponentene i nyttefunksjonen.

Modell iii viser at å ha norsk som morsmål vil ha en signifikant negativ effekt på utdanningsutfallet. Resultatene i Tabell 13 indikere at dette samsvarer med den positive, men ikke-signifikante, effekten for individer med innvandringsbakgrunn. Det er mulig dette kan forklares med kulturelle forskjeller, dvs. en del av de med innvandrerbakgrunn vil miste ansikt overfor familie dersom de avslutter studiet. Denne kulturen finner man ikke i like sterk grad hos norske studenter. Det å avslutte et studium anses ikke som noe tap av ansikt overfor venner og familie. I enkelte sammenhenger kan det tvert imot virke som noen av studentene har tatt et reflektert valg.

Modell iii viser at dersom studenten går 2-årig master eller profesjon vil det minske sannsynligheten for at de fortsetter studiet. Det kan tenkes at de som går 2-årig master har en lavere terskel for å avslutte studie da de allerede har en bachelorutdanning fra før. Sammenligner vi bachelor- og profesjonsstudium kan det tenkes at det er lettere for noen som går bachelor å fortsette, siden de ikke har noen utdanning på høyere nivå og det blir dermed viktig å fullføre. På en annen side kommer det fram i rapport til Andresen og Lervåg (2022) at det er høyest antall bytter og frafall for studenter som studerer på bachelor- og 2-årig masternivå, og tilsvarende lavere for profesjon og 5-årige masterprogram. Dette kan for eksempel begrunnes i at de som starter på profesjon eller 5-årige masterprogram har i større grad overveid valget sitt da det er en såpass lang utdanning, og dermed har mer motivasjon for å fortsette. De to motstridende resultatene kan skyldes at i vår modell er det ikke tatt hensyn til bytte av studium eller overføringer. Dette gjør at vår avhengige variabel fanger opp effekten av de som skal bytte studium på lik linje med de som planlegger å droppe ut.

Modell iii indikerer at utdanninger innenfor informasjonsteknologi og ingeniør har en negativ effekt på utdanningsutfallet. Bakgrunnen for dette kan være at studentenes forventninger til studiet ikke samsvarer med studiets innhold og oppbygning. Samtidig viser Andresen og Lervåg sin rapport (2022) at i disse fagfeltene er det færrest studenter som slutter.

I samme rapport kommer det også frem at utdanninger innenfor bedriftsøkonomiske og administrative fag vil ha negativ effekt på utdanningsutfall, noe som stemmer overens med vår analyse. Mulige årsaker til den negative effekten kan være at studenter innenfor økonomiske og administrative fag velger denne retningen fordi de er usikre på hva de vil. De velger dermed denne fagretningen fordi den er anvendbar innenfor mange bedrifter og organisasjoner. Det kan derfor være at motivasjonsgrunnet for det faglige innholdet er mindre enn hos andre faglige retninger. Imidlertid viser modellen at fagområdet innenfor lærerutdanning og pedagogikk vil ha en større sannsynlighet for frafall, til tross for at dette er en yrkesrettet utdanning.

I modell iv ble det lagt på et kvadratisk ledd. Dette var for å se på om effekten av arbeidstimer er avtakende, noe resultatene kunne bekrefte. Siden det er en konkav modell, vil denne modellen ha et toppunkt. Dette stemmer overens med tidligere litteratur, hvor de finner en terskel for hvor mange timer studenter kan jobbe før det går ut over utdanningsutfallet. Dersom *Arbeidstid* og *Arbeidstid_sqrd* hadde vært signifikant, kunne den optimale arbeidstiden blitt regnet ut. Siden dette ikke er tilfellet, er det ikke relevant å regne ut dette toppunktet. Determinasjonskoeffisienten, R^2 , i denne modellen økte ikke ved å legge til et kvadratisk ledd.

I kapittel 2.3 om samfunnsøkonomisk teori viste vi til hvilke privatøkonomiske og samfunnsøkonomiske kostnader og gevinster høyere utdanning kan bidra til. I artikkelen til Erling Brandt (2022) kommer det frem at et høyt utdannet samfunn bidrar til økt produktivitet og økonomisk vekst for samfunnet. Fra et samfunnsøkonomisk ståsted vil det være hensiktsmessig å redusere frafallet ved høyere utdanning. Ved å kartlegge årsaker til frafall i høyere utdanning, kan myndighetene innføre tiltak med den hensikten å senke frafallet, og på den måten øke den samfunnsøkonomiske avkastningen.

Det antas at modell i hadde utelatt variabelproblem. Ved å kontrollere for flere kontrollvariabler vil sannsynligheten for at dette er tilfellet bli redusert. Dette vil dog ikke sørge for at det ikke er utelatt variabelproblem fordi ikke alle variabler fra datasettet er lagt til. Dette kan forklare hvorfor R^2 er så lav, fordi det kan være andre variabler som forklarer studentenes frafall fra studie som ikke er tatt med i modellene. Flere rapporter og analyser, blant annet Hovdhaugen (2019), trekker fram at karakterer fra videregående skole er en betydningsfull faktor for utdanningsutfallet. Dette er ikke inkludert i våre modeller, noe som kan påvirke til en lav R^2 .

Et annet mulig avvik kan være at våre analyser ikke skiller mellom de som planlegger å avslutte studie og ikke fortsette, og de som avslutter studie for å bytte til et annet studium. Fra Andresen og Lervåg sin rapport (Andresen & Lervåg, 2022) observeres det at hver fjerde student som begynte på en gradsutdanning i 2012 hadde et sektorfravær fra den aktuelle utdanningen. Sektorfravær omhandler fravær fra høyere utdanning i ett eller flere år, og bytter, som innebærer bytte av studieprogram eller utdanningsinstitusjon. Dette betyr at den avhengige variabelen fanger opp effekten av de som skal bytte studium, noe som vil gi avvik i resultatene.

5.4 Mulige utvidelser

Mulige utvidelser for analysen kan være å legge til et interaksjonsledd. Med et interaksjonsledd kunne det blitt observert om effekten av arbeidstid påvirkes av andre variabler. Her kunne det for eksempel blitt sett på interaksjonsledd mellom *Arbeidstid* og kjønn, som kan fortelle om effekten av arbeidstiden vil være forskjellig for kvinner og menn. Et annet mulig interaksjonsledd vil være å se om det er noen forskjell på studenter med foreldre med høyere utdanning og de studentene som ikke har foreldre med høy utdanning. En videre mulig utvidelse er å se på en modell hvor *Arbeidstid* og *Arbeidstid_sqrd* har et interaksjonsledd med *kvinner*. Ved en slik modell vil vi kunne se om en optimal arbeidstid vil være ulik for kvinner og menn.

Modellen kunne også ha blitt utvidet ved å ta inn flere kontrollvariabler og muligens fått en høyere R^2 . Dersom datasettet hadde inkludert flere land, kunne analysen blitt gjennomført for ytterligere land. Da ville det vært mulig å sett om effekten av *Arbeidstid* ville vært forskjellig for norske og utenlandske studenter.

En siste mulig utvidelse kan være eventuelle svakheter ved OLS-modellene. Modellen sikrer ikke at predikerte sannsynligheter ligger mellom 0 og 1, noe som i utgangspunktet taler for å bruke en logit- eller probit-spesifikasjon. Dersom vi hadde brukt logit- eller probit-spesifikasjon ville det muligens gi en høyere R^2 , og dermed en større forklaringskraft (Wooldridge, 2019. s.240).

5.5 Oppsummering

I dette kapittelet presenteres resultatene fra regresjonsanalysene. Vi startet analysen med en enkel lineær modell. Deretter utvides modellen ved å legge til flere kontrollvariabler, først for de sosiodemografiske kjennetegnene og deretter for utdanningsbakgrunnen til studenten. I den siste modellen inkluderes et kvadrert ledd, hvor det ble observert at interessevariabelen *Arbeidstid* er positiv, men i en avtagende grad. Deretter diskuteres resultatene og foreslås mulige utvidelser.

6 Konklusjon

I denne oppgaven har følgende problemstilling blitt analysert: “Er det sammenheng mellom studentenes frafall fra studier og lønnet arbeid?”. Resultatene viser at arbeidstid har en liten signifikans på om studentene velger å fortsette studiet, men en kan ikke konkludere med at effekten av arbeidstid ikke er til stede. Vi finner at det er de sosiodemografiske kjennetegnene til individet som i større grad har signifikant effekt på om de velger å fortsette påbegynt utdanning. Resultatene i denne oppgaven er til en viss grad sammenfallende med tidligere forskning. På tross av relativt liten forklaringskraft i modellene er det trolig at våre resultater er valide, men i videre forskning er det relevant å inkludere variabler som i større grad kan forklare hvorfor studenter egentlig slutter.

7 Referanser

- Andresen, S. M. H. og Lervåg, M. L. (2022). *Frafall og bytter i universitets- og høgskoleutdanning*. (SSB-rapport 2022/6). Oslo: SSB. Tilgjengelig fra: https://www.ssb.no/utdanning/hoyere-utdanning/artikler/frafall-og-bytter-i-universitets-og-hogskoleutdanning/_attachment/inline/c2b8fce1-6f81-480e-82fa-3452bddd1916:623025b29120ba5200bda3a2cdc8d17147375682/RAPP2022-06.pdf (Hentet: 09. mai 2022)
- Barth, E. (2005) Den samfunnsmessige avkastning av utdanning, *Utdanning 2005 – deltakelse og kompetanse*, Statistiske analyser 74, s.168-190. Tilgjengelig fra: https://www.ssb.no/a/publikasjoner/pdf/sa74/kap-8.pdf?fbclid=IwAR1jfEnd8tVjRQakVVMjYkVhTJpPLXhKY5o2tpWY09MixT3F6z_Am8sh_KI
- Forskrift til utdanningsstøtte (2020) *Behovsprøving av stipend* (utdanningsstøtteloven). Tilgjengelig fra: <https://lanekassen.no/nb-NO/regelverk/tildeling/del3-kapittel12/slik-fungerer-omgjoring/> (Hentet: 27. april 2022)
- Hovdhaugen, E. (2015) *Journal of Education and Work, Working while studying: the impact of term-time employment on dropout rates*, 28(6), s. 17-18. Doi: 10.1080/13639080.2013.869311
- Hovdhaugen, E. (2019) *Årsaker til frafall i høyere utdanning*, (21030) Oslo: NIFU. Tilgjengelig fra: <https://nifu.brage.unit.no/nifu-xmlui/bitstream/handle/11250/2593810/NIFUarbeidsnotat2019-3.pdf?sequence=1&isAllowed=y> (Hentet: 09.mai 2022)
- Jæger, M. M. (2007). Economic and Social Returns to Educational Choices: Extending the Utility Function, *Rationality and Society*, bind 19, nr. 4, s.451-483. <https://journals.sagepub.com/doi/pdf/10.1177/1043463107083739> Doi: 10.1177/1043463107083739
- Katanasho, M. B. (2022). Forsker: så mye bør du jobbe deltid ved siden av studiene, *Universitas.no*. Tilgjengelig fra: https://universitas.no/nyheter/69092/forsker-sa-mye-bor-du-jobbe-deltid-ved-siden-av-st/?fbclid=IwAR2R3hM_iYlyeXksKB8Upnnbj_jZr3WSY4a5dt-Mirxh92q-Tft__rqOII.

- Moulin, S. Doray, P. Laplante, B. og Street, M, C.(2013) *Journal of Education and Work, Work intensity and non-completion of university: longitudinal approach and causal inference*,26(3), s. 352-353. Doi: 10.1080/13639080.2011.653554
- OECD (2019), *Education at a Glance 2019: OECD Indicators*, OECD Publishing, Paris, <https://doi.org/10.1787/f8d7880d-en>.
- Statistics Norway. (2022). *Eurostudent VI, 2016* [Data set]. Sikt - Norwegian Agency for Shared Services in Education and Research. <https://doi.org/10.18712/NSD-NSD2560-V2>
- St.meld.13(2018-2019). (2019) *Muligheter for alle – Fordeling og sosial bærekraft*. Oslo: Finansdepartementet. Tilgjengelig fra: <https://www.regjeringen.no/no/dokumenter/meld.-st.-13-20182019/id2630508/?ch=4> (Hentet 09. mai 2022)
- Utdanningsstøtteleven (2005) Lov om utdanningsstøtte. Tilgjengelig fra: <https://lovdata.no/dokument/NL/lov/2005-06-03-37> (Hentet: 28. april 2022)
- Wooldridge, J.M. (2019) *Introductory Econometrics – A Modern Approach*. 7. Utgave. Boston, USA: Cengage.

Vedlegg 1: Korrelasjonsmatrise

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	
(1) y1	1.000																									
(2) Arbeidstid	-0.053	1.000																								
(3) Arbeidstid_sqrd	-0.066	0.947	1.000																							
(4) alder_u22	0.137	-0.231	-0.217	1.000																						
(5) alder22_24	0.059	-0.191	-0.214	-0.351	1.000																					
(6) alder25_29	-0.080	-0.028	-0.053	-0.290	-0.378	1.000																				
(7) mig1	0.014	-0.017	-0.013	-0.093	-0.082	0.068	1.000																			
(8) mig2	0.024	-0.024	-0.029	0.039	0.040	-0.004	-0.105	1.000																		
(9) female	0.001	0.075	0.046	0.033	-0.043	-0.047	-0.021	-0.017	1.000																	
(10) velstand	0.005	-0.091	-0.095	0.082	0.084	-0.023	-0.088	-0.011	-0.045	1.000																
(11) middelsvelstand	-0.008	0.068	0.071	-0.048	-0.042	0.009	-0.015	-0.007	0.055	-0.808	1.000															
(12) pared	0.013	-0.090	-0.095	0.060	0.059	0.029	-0.054	-0.024	-0.026	0.194	-0.059	1.000														
(13) barn	-0.073	0.392	0.422	-0.252	-0.303	-0.123	0.059	-0.066	0.117	-0.106	0.056	-0.149	1.000													
(14) morsmal	-0.015	0.019	0.014	0.095	0.076	-0.062	-0.810	0.004	0.017	0.089	0.007	0.082	-0.059	1.000												
(15) bachelor	0.258	-0.218	-0.258	0.201	0.149	-0.106	-0.060	0.030	0.008	0.009	-0.002	-0.042	-0.196	0.060	1.000											
(16) master	-0.302	0.040	0.069	-0.161	-0.038	0.115	0.073	-0.019	-0.074	0.020	-0.016	0.041	0.045	-0.075	-0.615	1.000										
(17) profesjon	-0.049	-0.090	-0.101	-0.006	0.034	0.068	-0.038	0.028	0.045	0.034	-0.020	0.093	-0.084	0.044	-0.279	-0.161	1.000									
(18) human	0.042	-0.008	-0.012	-0.009	-0.006	0.028	0.022	0.006	0.004	-0.034	0.015	0.010	-0.016	-0.031	-0.023	0.013	-0.039	1.000								
(19) samf	0.025	0.005	-0.005	0.010	-0.013	0.016	0.034	0.003	0.028	0.003	-0.008	0.020	-0.044	-0.030	-0.000	0.028	-0.025	-0.096	1.000							
(20) helse	0.025	0.015	-0.010	-0.027	-0.035	-0.007	-0.019	-0.008	0.186	-0.032	0.034	-0.021	0.093	0.024	0.064	-0.147	0.180	-0.146	-0.153	1.000						
(21) it	-0.040	-0.070	-0.056	0.009	0.042	-0.001	-0.000	0.021	-0.175	0.032	-0.038	0.006	-0.074	-0.003	0.028	0.041	-0.053	-0.064	-0.067	-0.103	1.000					
(22) naturvit	0.042	-0.094	-0.079	0.046	0.014	0.003	0.029	0.010	-0.058	-0.005	0.011	0.051	-0.077	-0.027	-0.043	0.067	-0.008	-0.085	-0.089	-0.136	-0.060	1.000				
(23) laerer	-0.002	0.111	0.106	-0.059	-0.045	-0.007	-0.038	-0.038	0.118	-0.052	0.055	-0.025	0.128	0.042	-0.026	-0.071	-0.058	-0.131	-0.137	-0.209	-0.092	-0.122	1.000			
(24) ingeniør	-0.087	-0.160	-0.129	0.027	0.068	0.016	-0.010	0.007	-0.205	0.038	-0.038	0.018	-0.109	0.004	0.001	0.105	-0.045	-0.112	-0.118	-0.179	-0.079	-0.104	-0.160	1.000		
(25) bedok	-0.022	0.102	0.090	0.014	0.013	-0.021	-0.014	0.009	-0.017	0.071	-0.051	-0.010	0.004	0.011	-0.002	0.051	-0.004	-0.147	-0.155	-0.236	-0.104	-0.137	-0.211	-0.181	1.000	

