

Eksisterer det et lønnsgap mellom innvandrere og etniske nordmenn?

Hovedoppgave i Bachelor samfunnsøkonomi

Veileder: Bjarne Strøm

Mai 2019

-

Eksisterer det et lønnsgap mellom innvandrere og etniske nordmenn?

Bacheloroppgave i Samfunnsøkonomi
Veileder: Bjarne Strøm
Mai 2019

Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
Institutt for samfunnsøkonomi

 **NTNU**
Norwegian University of
Science and Technology

Contents

1	Innledning	2
2	Teori	3
2.1	Regresjonsanalyse og <i>ordinary least squares</i>	3
2.1.1	Regresjon med to variabler	3
2.1.2	OLS	4
2.1.3	Føyningsmål	5
2.1.4	Multippel regresjon	6
2.2	Tidligere studier	7
3	Data	9
3.1	Datasett	9
3.2	Variabler	10
3.3	Deskriptiv statistikk for kontrollvariabler	11
4	Regresjonsanalyse	13
4.1	Modeller	13
4.2	Modeller	13
4.3	Resultat	14
4.3.1	Eksisterer det et lønnsgap	14
4.3.2	Kan lønnsgap forklarerer med forskjell i arbeidserfaring? . . .	15
4.3.3	Kan lønnsgap forklarerer med forskjell i utdanningsnivå? . . .	16
4.3.4	Kan lønnsgap forklarerer med forskjell i tallferdigheter?	16
4.3.5	Kan lønnsgap forklarerer med forskjell i andre ferdigheter? . .	16
4.4	Lønnsavkastning	17
4.5	Begrensninger	18
5	Oppsummering	21
6	referanser	22

1 Innledning

* Innvandring er i dag et stort tema i de nordiske landene, og partier med innvandringsfiendtlig retorikk vokser.

Derfor er integrering et viktig tema. Derfor vil jeg se på innvandreres lønn målt mot majoritetsbefolkningen i Norge.

- Har innvandrere samme lønn som majoritetsbefolkningen?
- Får innvandrere samme lønn når vi kontrollerer for utdanningsnivå, og ferdigheter?
- Har innvandrere samme avkastning på utdanning og ferdigheter som majoritetsbefolkningen?

Oppgaven vil starte med teori der vårt statistiske grunnlag blir utledet, samt supplerende teori om innvandring og lønn. Deretter vil datasett presenteres. Så følger en analyse av datasettet før hovedpunktene vil bli oppsumert.

2 Teori

2.1 Regresjonsanalyse og *ordinary least squares*

2.1.1 Regresjon med to variabler

I vår empiriske analyse av lønnsforskjeller mellom innvandrere og majoritetsbefolkningen vil vi bruke metoden *ordinary least squares*, *OLS*, som er en form for regresjon. Regresjonsanalyse tar i sin enkleste form utgangspunkt i observasjoner av to variabler, vanligvis angitt X og Y , og antar at den ene variabelen er avhengig av den andre variabelen, og at forholdet mellom dem er lineært. Deretter vil vi forsøke å se hvor mye variablene avhenger av hverandre. Vi starter med å se på forholdet mellom gitt X og Y , der Y er avhengig variabel, og X og Y er mengder med observasjoner, der hver observasjon er et reelt tall.

$$E(Y) = \alpha + \beta X,$$

altså at gitt X angir en verdi for forventningsverdien til Y . α og β er hhv. punktet der linjen produsert av ligningen krysser y-aksen og stigningshøyden. Siden forventningsverdien ikke alltid stemmer helt overens med virkeligheten vil Y være på formen

$$Y = E(Y) + \epsilon$$

Dette kan vi sette inn for $E(Y)$ i 2.1.1, som gir

$$Y = \alpha + \beta X + \epsilon, \tag{2.1}$$

her er ϵ estimeringsfeilen, ofte kalt residualleddet, og tar høyde for at X og koeffisientene α og β ikke estimerer Y eksakt. Regresjonsanalyse blir brukt til å se på forholdet mellom mengder av observasjoner $X = \{x_1, x_2, \dots, x_n\}$ og $Y = \{y_1, y_2, \dots, y_n\}$. Vi benevner størrelsen på mengden n og i angir en gitt observasjon i mengden. Da har vi for gitt i

$$Y_i = \alpha + \beta X_i + \epsilon_i, \tag{2.2}$$

men ofte vil vi se på forholdet mellom variablene X og Y for alle tall, og ikke bare for gitte observasjoner. Derfor må vi estimere forholdet mellom Y og X , ved å velge estimerer for koeffisientene α og β , vanligvis betegnet a og b . Hvis vi antar at estimert verdi av Y til å være \hat{Y} , vil vi ha $\hat{Y} = a + bX$, da vil estimeringsfeilen for hver observasjon, ϵ_i , være differansen mellom estimert verdi og faktisk verdi av Y_i .

$$\epsilon_i = Y_i - \hat{Y}_i \tag{2.3}$$

2.1.2 OLS

For å bruke OLS er det fem forutsetninger som må være oppfylt:

1. Ikke-stokastiske X ,
2. $E(\epsilon_i) = 0$,
3. $Var(\epsilon_i) = E(\epsilon_i^2) = \sigma^2$,
4. $cov(\epsilon_i, \epsilon_j) = 0$,
5. $\epsilon_i \sim N(0, \sigma^2)$,

OLS tar utgangspunkt i enkel regresjonsanalyse. Prinsippet er at man skal minimere summen av residualleddet, e_i kvadrert. I enkel regresjon har vi da;

$$S = \sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - a - bX_i)^2 \quad (2.4)$$

Denne summen kalles *SSR*, *square sum of residuals*. dermed blir problemet å velge a og b slik at S blir minst mulig. Vi kan finne a og b ved å derivere.

$$\frac{\partial S}{\partial a} = -2\sum (Y_i - a - b - X_i) = 0, \quad (2.5)$$

$$\frac{\partial S}{\partial b} = -2\sum X_i(Y_i - a - b - X_i) = 0, \quad (2.6)$$

Ut ifra 2.5 finner vi a :

$$-2\sum (Y_i - a - b - X_i) = 0 \Rightarrow \sum Y_i = na + b\sum X_i \Rightarrow \frac{\sum Y_i}{n} = a + b\frac{\sum X_i}{n} \quad (2.7)$$

$$\Rightarrow a = \bar{Y} - b\bar{X} \quad (2.8)$$

der

$$\frac{\sum Y_i}{n} = \bar{Y},$$
$$\frac{\sum X_i}{n} = \bar{X},$$

altså er a estimert ved gjennomsnittsverdien til mengdene X og Y , samt koeffisienten b . b finner vi ved 2.6

$$-2\sum X_i(Y_i - a - b - X_i) = 0 \Rightarrow \sum Y_i X_i = a\sum X_i + b\sum X_i^2$$

Vi setter så inn for a

$$\Sigma Y_i X_i = a \Sigma X_i + b \Sigma X_i^2 \Rightarrow \Sigma Y_i X_i = \frac{\Sigma X_i \Sigma_i}{n} + b \left[\Sigma X_i^2 - \frac{(\Sigma X_i)^2}{n} \right]$$

$$b = \frac{\Sigma X_i Y_i - (1/n) \Sigma X_i \Sigma Y_i}{\Sigma X_i^2 - (1/n) (\Sigma X_i)^2}$$

Det kan vises at

$$\Sigma (X_i - \bar{X})^2 = \Sigma X_i^2 - \frac{(\Sigma X_i)^2}{n}$$

og at

$$\Sigma (X_i - \bar{X})(Y_i - \bar{Y}) = \Sigma X_i Y_i - \frac{\Sigma X_i \Sigma Y_i}{n},$$

dermed har vi

$$b = \frac{\Sigma (X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma (X_i - \bar{X})^2}, \quad (2.9)$$

Når estimatorene b utledes på denne måten kan det vises at de er lineære estimatorer av β . Dette betyr at $E(b) = \beta$. Siden vi har forutsetningen om at ϵ_i er normalfordelt er Y_i også normalfordelt. Det følger da at b er normalfordelt, siden b er en lineær estimator. Det samme med a .

2.1.3 Føyningsmål

Føyningsmål ser på hvor mye av variasjonen i Y som kan forklares av vår modell. For å beregne dette trenger vi tre ligninger.

$$SST = \Sigma (Y_i - \bar{Y})^2 \quad (2.10)$$

$$SSE = \Sigma (\hat{Y}_i - \bar{Y})^2 \quad (2.11)$$

$$SSR = \Sigma \epsilon_i^2 \quad (2.12)$$

SSR er den uforklarte variasjonen, SSE er en forklarte variasjonen, mens SST er den totale variasjonen, som er slik at

$$SST = SSE + SSR,$$

Ut ifra dette definerer vi modellens forklaringskraft, R^2 til å være

$$R^2 = \frac{SSE}{SST} = \frac{SST - SSR}{SST} = 1 - \frac{SSR}{SST} \quad (2.13)$$

Det er trivielt at $0 \leq R^2 \leq 1$. Hvis $R^2 = 0$ forklarer modellen ingenting av variasjonen i Y , mens modellen forklarer alt av variasjon i Y når R^2 .

2.1.4 Multippel regresjon

Til nå har vi sett på regresjon med to variabler. I analysen senere vil vi dog bruke modeller på formen.

$$Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + \epsilon, \quad (2.14)$$

Dette betyr at hver enkelt observasjon av Y , Y_i , vil kunne uttrykkes på formen

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \epsilon_i, \quad (2.15)$$

For multippel regresjon trenger vi en ekstra antagelse om vår modell, nemlig at ingen av variablene X_{ji} er en lineær kombinasjon av noen av de andre variablene. Da vil estimatorene vi finner være forventningsrette og effisiente.

På samme måte som i tilfellet med to variabler, vil OLS gå ut på å finne estimerer b_1, b_2, \dots, b_k for koeffisientene β_j , slik at vi minimerer summen

$$S = \sum \epsilon_i^2 \quad (2.16)$$

hvor $\epsilon_i = Y_i - \hat{Y}_i$ som tidligere. Men tidligere hadde vi kun to avhengige variabler, mens det nå er k variabler. Dette medfører at vi får k førsteordensbetingelser når vi deriverer 2.15 med hensyn på alle b_j , på formen:

$$\frac{\partial S}{\partial b_j} = 2 \sum X_{ji} (Y_i - b_1 - b_2 X_{2i} - \dots - b_k X_{ki}), \quad X_{1i} = 0,$$

Analysen senere vil bygge på estimatorene b_j utledet på denne måten. Det kan vises at variansen til disse estimatorene er

$$\sigma_{b_j}^2 = \frac{\sigma_i^2}{\sum X_{ji}^2 (1 - r)} \quad (2.17)$$

der r er et ledd for å fange opp korrelasjonen mellom variablene X_{ji} , og σ_i^2 variansen til ϵ_i . Siden vi sjelden kjenner variansen vil vi måtte bruke den forventningsrette estimatoren til σ_i^2 , s^2

$$s^2 = \frac{\sum \epsilon_i}{n - k} \quad (2.18)$$

der n er antall observasjoner og k er antall parametre. Det følger da at vi benytter variansen for de ulike estimatorene b utledet med s^2 , s_{b_j} . Senere vil vi teste hypoteser om estimatorene. Siden vi ikke kjenner variansen, men estimatorene er normalfordelt, følger fra at ϵ_i er normalfordelt, vil testobservatoren i tilfellene hvor

vi tester hypoteser på én og én variabel, være t-fordelt med $n - k$ frihetsgrader. Da vil testobservatoren være på formen;

$$TS = \frac{b_j - \beta_j}{\sigma_{b_j}} \sim t_{n-k} \quad (2.19)$$

R^2 vil utledes på lik måte som i den enkle modellen.

Når vi vil teste hypoteser om flere av variablene samtidig, vil vi benytte F-test. F. eks. skal teste en hypotese om at $\beta_1 = \beta_2 = \beta_3 = 0$ i 2.15. Vil da først finne estimatorene for modellen, deretter formulere modellen der restriksjonen over, $\beta_1 = \beta_2 = \beta_3 = 0$. holder. La SSR_u og SSR betegne henholdsvis modell uten og med restriksjon. Da vil testobservatoren for en slik hypotesetest være

$$TS = \frac{(SSR_u - SSR)/h}{SSR_u/(n - k)} \sim F_{h,n-k}, \quad (2.20)$$

der h betegner antall variabler vi tester, i dette tilfellet $h = 3$, n er antall observasjoner og k er antall variabler totalt. Fra tidligere husker vi at $R^2 = 1 - \frac{SSR}{SST}$. Setter vi inn dette får vi

$$TS = \frac{(R_u^2 - R^2)/h}{R^2/(n - k)}, \quad (2.21)$$

2.2 Tidligere studier

Innvandrere er i vår analyse en som er født i utlandet med minst en utenlandsk forelder.

SSB skrev i 2017 at innvandrere har lavere lønn enn majoritetsbefolkningen i Norge. De fant at i snitt tjente innvandrere 5300kr mindre i måneden enn etniske nordmenn, basert på tall fra lønnstatistikken innrapportert til A-ordningen i september 2016, hvor innvandrere er skilt ut som egen gruppe, deretter i fødeland. A-ordningen er ”en samordnet måte for arbeidsgivere å rapportere opplysninger om inntekt og ansatte til NAV, SSB og Skatteetaten” (Skatteetaten, 2019). De pekte på forskjeller i kompetanse som en viktig bidragsyter til dette. Samtidig er det store forskjeller mellom innvandreregrupper. Afrikanske innvandrere tjener vesentlig mindre enn innvandrere fra vest-Europa.

SSB konkluderer også med at innvandrere og etniske nordmenn systematisk er ansatt i forskjellige yrker, og at lønnsforskjellene minker når det kontrolleres for sektor. I enkelte sektorer var det ingen forskjell i lønn mellom innvandrere og etniske nordmenn, mens i andre, som snekker, tjente innvandrere i snitt 94% av etniske nordmenn.

Svenske Institutet för arbetsmarknads- och utbildningspolitisk utvärdering, publicerade i 2013 en rapport med titel *Invandrade med svensk utbildning har lägre lön än infödda*. Her skriver forfatterne ”personer med svensk bakgrund har högre lön än utrikes födda, även om dessa kommit till Sverige som barn och har lika hög utbildning”.

3 Data

I denne delen vil datasettet og forklaringsvariabler presenteres og defineres.

3.1 Datasett

Datasettet vi bruker er hentet fra OECDs PIACC undersøkelse. Denne så på den voksne befolkningen i 24 land, og kartla deres ferdigheter innen tallforståelse, problemløsning og lesing. I tillegg måtte respondentene svare på en rekke spørsmål som; lønn, arbeidserfaring osv.

I vår analyse vil vi se på mennesker over 25 år. Dette for å hindre seleksjonsproblemer ved å ta med folk som fortsatt er under utdanning i analysen. Et sentralt begrep i analysen vil være innvandrere. I dette datasettet er innvandrere definert som en person som er født i utlandet, og har minst en forelder som også er født i utlandet. Mulige mangler i analysen pga. at det ikke inkluderes i datasettet er foreldres lønnsnivå, vi vet ikke om innvandreren er flyktning eller hvilket land innvandrerne kommer fra. Vi vet at lønnsnivået til foreldre har en korrelasjon med lønnsnivået til sine barn. Det hadde også vært interessant å se på lønnsnivå i forskjellige innvandringsgrupper, gruppert etter fødselsland. Det er nærliggende å tenke at høytudannede flyktninger fra ikke-vestlige land ikke får arbeid eller lønn som tilsvarer f. eks. europeere med samme utdanningsnivå.

3.2 Variabler

- Avhengig variabel:
 - *earnhrppp* - kjøpekraftsjustert timelønn i dollar,\$,
 - * $lwage = \log(earnhrppp)$
- Sentral forklaringsvariabel:
 - *migrant_first* - dummy-variabel, 1 hvis innvandrer, 0 ellers,
- Kontrollvariabler:
 - *yrqual* - antall år utdanning,
 - *c_q09_c* - antall år arbeidserfaring,
 - * $c_{q09_c_sq} = (c_{q09_c})^2$
 - *litscore1* - leseforståelse,
 - *numscore1* - tallferdigheter,
 - *pslscore1* - problemløsingsscore,

3.3 Deskriptiv statistikk for kontrollvariabler

Tabell 1: Deskriptiv statistikk for majoritetsbefolkningen,
migrant_first = 0

Variable	Obs	Mean	Std. Dev.	Min	Max
earnhrppp	2,594	26.0849	8.488364	5.892016	64.50208
yrsqual	3,394	14.6594	2.347391	7	21
c_q09_c	3,375	22.8883	11.35843	0	47
litscore1	3,395	.1346665	.9019993	-4.214909	2.818439
numscore1	3,395	.1595228	.8880019	-3.513036	2.571487
pslscore1	3,001	-.0155342	.9766591	-3.594971	3.20957

Deskriptiv statistikk for innvandrere, *migrant_first = 1*

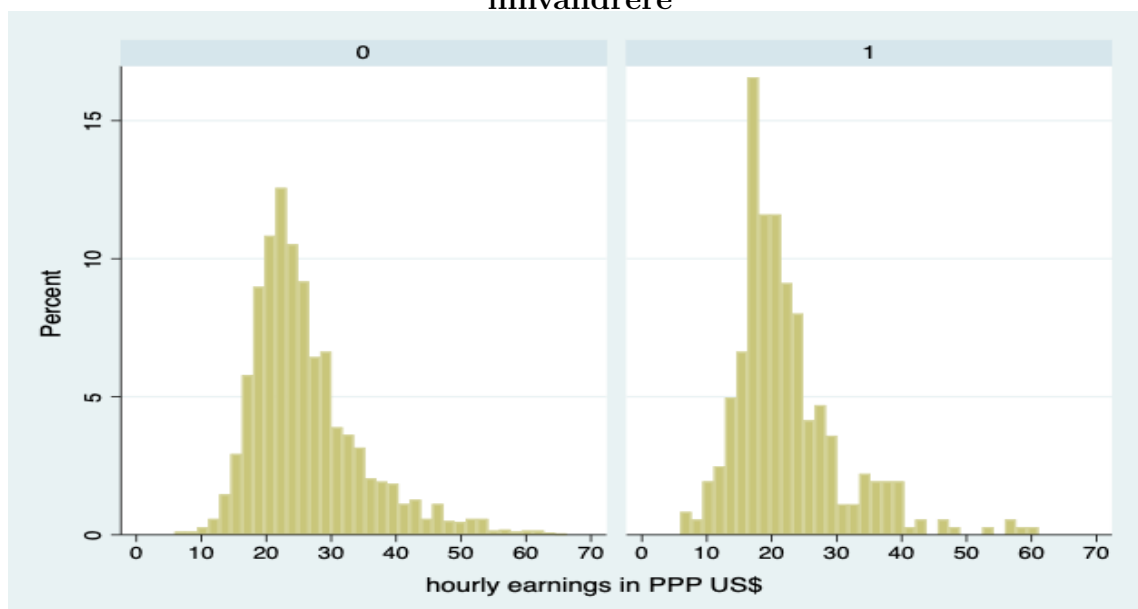
Variable	Obs	Mean	Std. Dev.	Min	Max
earnhrppp	362	22.24828	8.361452	7.040093	60.65311
yrsqual	498	14.79518	3.088601	7	21
c_q09_c	480	16.05625	10.46951	0	47
litscore1	498	-.7327305	1.355738	-4.652344	2.016985
numscore1	498	-.7995764	1.36891	-4.327034	1.801197
pslscore1	344	-.5104061	1.19023	-5.512477	2.321082

Kommentarer:

- Majoritetsbefolkningen tjener i gjennomsnitt i overkant av 3,7\$ mer enn innvandrere, men standardavviket er omtrent det samme.
- Antall år utdanning er så og si lik mellom de to gruppene, men standardavviket hos majoritetsbefolkningen er litt mindre.
- Majoritetsbefolkningen har i snitt i underkant av sju år mer arbeidserfaring.
- Innvandrere scorer dårligere på ferdighetstester enn majoritetsbefolkningen, men det er også et betydelig større standardavvik.

- Mulig forklaring på lønns-gap kan være forskjeller i arbeidserfaring og ferdigheter.

Histogram med lønns-spredning for hhv. majoritetsbefolkning og innvandrere



Kommentar: Tydelig at lønnen til innvandrere er mer sentrert rundt 20\$, mens det er jevnere spredning blant majoritetsbefolkningen.

4 Regresjonsanalyse

Her vil regresjonsmodeller for analyse og resultatet av analysen presenteres. Resultatet vil brukes til å svare på problemstillingene innledningen.

4.1 Modeller

Siden vi skal se på hvordan lønn varierer med forskjellige variabler vil *earnhrppp* være vår avhengige variabel i alle modeller, men vi vil benytte en logaritmisk lønnsfunksjon $lwage - \log(earnhrppp)$, dette er fordi logaritmisk lønnsfunksjon kan tolkes som tilnærmet prosentvis lønns-gap.

Vi skal sammenligne majoritetsbefolkning og innvandrere, som betyr at *migrant_first* vil være en del av alle modeller. Vi skal kontrollere for forskjellige faktorer, disse vil introduseres en og en.

4.2 Modeller

1. $lwage_i = \beta_1 + \beta_2 migrant_first + \epsilon_i$,
2. $lwage_i = \beta_1 + \beta_2 migrant_first + \beta_3 c_q09_c + \beta_4 c_q09_c_sq + \epsilon_i$,
3. $lwage_i = \beta_1 + \beta_2 migrant_first + \beta_3 c_q09_c + \beta_4 c_q09_c_sq + \beta_5 yrsqual + \epsilon_i$,
4. $lwage_i = \beta_1 + \beta_2 migrant_first + \beta_3 c_q09_c + \beta_4 c_q09_c_sq + \beta_5 yrsqual + \beta_6 numscore1 + \epsilon_i$,
5. $lwage_i = \beta_1 + \beta_2 migrant_first + \beta_3 c_q09_c + \beta_4 c_q09_c_sq + \beta_5 yrsqual + \beta_6 numscore1 + \beta_7 litscore1 + \beta_8 pslcore1 + \epsilon_i$,

Modell (1) angir lønnsforskjell mellom innvandrere og majoritetsbefolkning når vi ikke kontrollerer for andre faktorer.

Modell (2) kontrollerer i tillegg for arbeidserfaring, samt arbeidserfaring kvadrert, i tilfelle effekten av arbeidserfaring ikke er lineær.

Modell (3) kontrollerer i tillegg for utdanningsnivå.

Modell (4) tar med tallferdigheter

Modell (5) kontrollerer for problemløsning og leseferdigheter i tillegg.

4.3 Resultat

Tabell 3: Resultater av regresjonanalyse

VARIABLES	(1) A lwage	(2) B lwage	(3) C lwage	(4) D lwage	(5) E lwage
migrant_first	-0.17305 (0.01733)	-0.13925 (0.01750)	-0.13459 (0.01606)	-0.05712 (0.01670)	-0.04796 (0.01809)
c_q09_c		0.01909 (0.00210)	0.01968 (0.00193)	0.01822 (0.00188)	0.01924 (0.00202)
c_q09_c_sq		-0.00035 (0.00004)	-0.00031 (0.00004)	-0.00026 (0.00004)	-0.00026 (0.00004)
yrsqual			0.05321 (0.00226)	0.04041 (0.00241)	0.03984 (0.00265)
numscore1				0.08157 (0.00623)	0.08274 (0.01304)
litscore1					-0.04289 (0.01346)
pslscore1					0.04867 (0.00983)
Constant	3.21403 (0.00607)	3.00132 (0.02271)	2.16818 (0.04110)	2.34113 (0.04209)	2.33884 (0.04592)
Observations	2,956	2,954	2,953	2,953	2,672
R-squared	0.033	0.063	0.211	0.254	0.240

Standard errors in parentheses

Her er resultatet av regresjonsanalysen presentert i Tabell 3. Vi kan se at estimert lønnsgap varierer mellom 17.31% og 4.8%. I alle påfølgende tester vil vi bruke 5% signifikansnivå, $\alpha = 0.05$

4.3.1 Eksisterer det et lønnsgap

Først undersøker vi en hypotese om at det ikke eksisterer lønnsforskjell mellom innvandrere og majoritetsbefolkningen. Vi formulerer hypotesesten

$$H_0 : b_2 = 0, H_A : b_2 > 0,$$

velger et signifikansnivå på 5% $\alpha = 0.05$. Da har vi kritisk verdi $t = 1.960$ (selv om antall frihetsgrader varierer mellom 2954 og 2664, vil TS være lik, siden antall

frihetsgrader er så stor i alle modeller). Vi forkaster nullhypotesen om absoluttverdien av testobservatoren er større enn dette. Beregner testobservator som beskrevet i teorien. Da får vi

1. $|TS| = 9.98$
2. $|TS| = 7.96$
3. $|TS| = 8.38$
4. $|TS| = 3.42$
5. $|TS| = 2.65$

Vi ser at vi forkaster nullhypotesen. I alle modeller er det lønnsforskjell mellom innvandrere og majoritetsbefolkningen. Vil nå gå gjennom hver modell systematisk, og se på de forskjellige kontrollvariablene.

4.3.2 Kan lønnsgap forklares med forskjell i arbeidserfaring?

For å teste om lønnsgapet kan forklares med forskjeller i arbeidserfaring formulerer vi testen:

$$H_0 : \beta_3 = \beta_4 = 0, H_A : \beta_3 \neq 0 \cup \beta_4 \neq 0$$

Vi kjenner igjen dette som et klassisk tilfelle for å bruke F-test med hhv. 2 og 2950 frihetsgrader. Da har vi kritisk verdi $F_{2,2950} = 2.9957$. Fra stata får vi TS:

```
. test c_q09_c c_q09_c_sq
```

```
( 1)  c_q09_c = 0
( 2)  c_q09_c_sq = 0
```

```
F( 2, 2950) = 47.24
Prob > F = 0.0000
```

Vi ser altså at arbeidserfaring har innvirkning på lønna. Vi ser også at lønnsgapet blir redusert med rundt 3% når vi kontrollerer for erfaring, men at det fortsatt holder seg ganske stort.

4.3.3 Kan lønnsgap forklarerer med forskjell i utdanningsnivå?

Vi kan først se at estimert effekt av utdanning på lønna er 5.321%. Det vil si at ett års ekstra utdanning gir 5.321% høyere lønn. Sjekker hypotese om at utdanning har effekt på lønn. Tilsvarende som over, t-test. Vi får $TS = 23.52$ som er høyere enn vår kritiske verdi, 1.950. Utdanning har effekt på lønn. Dog ser vi at når vi kontrollerer for utdanning reduseres lønnsgapet kun med 0.5%.

4.3.4 Kan lønnsgap forklarerer med forskjell i tallferdigheter?

Vi tester først om tallferdigheter har påvirkning på lønna. Gjøres ved t-test som over. $TS = 13.1$, dette er større enn vår kritiske verdi, altså har tallferdigheter påvirkning på lønna. Vi ser også at lønnsgapet reduseres med nesten 8% når vi kontrollerer for tallferdigheter gitt utdanning og erfaring. Det betyr at forskjeller i tallferdigheter mest sannsynlig står for en del av lønnsgapet.

4.3.5 Kan lønnsgap forklarerer med forskjell i andre ferdigheter?

Vi ser at paradoksalt nok er leseferdigheter forbundet med en lønnsreduksjon på 4.289%. Samtidig er problemløsning forbundet med en lønnsøkning på 4.867%. Tester en hypotese om at disse ikke har innvirkning på lønna med F-test. Vi får $F_{3,2664} = 2.9957$. Fra stata:

```
. test litscore1 numscore1 pslscore1

( 1)  litscore1 = 0
( 2)  numscore1 = 0
( 3)  pslscore1 = 0

          F( 3, 2664) =    60.60
          Prob > F =    0.0000
```

Altså $TS = 60.60$ leseferdigheter og problemløsning har innvirkning på lønn. Utifra tabellen, kan det ut som om disse ferdighetene står for i underkant av 1% når de andre faktorene er kontrollert for.

4.4 Lønnsavkastning

Vi utvider nå modell 5 til å inkludere interaksjonsvariabler, variabler som ser på sammenhengen mellom variabler. På denne måten kan vi se om innvandrere har forskjellig avkastning på enkelte av kontrollvariablene.

Vi definerer $yrqual_mig = migrant_first \cdot (yrqual - \bar{yrqual})$, $num_mig = migrant_first \cdot (numscore1 - \bar{numscore1})$, $lit_mig = migrant_first \cdot (litscore1 - \bar{litscore1})$ og $psl_mig = migrant_first \cdot (pslcore1 - \bar{pslcore1})$. Der Vi benytter oss av avviket fra gjennomsnittsverdien til de forskjellige variablene. Dette er fordi hvis vi i stedet hadde benyttet oss av f. eks. $yrqual_mig = migrant_first \cdot (yrqual)$ ville vi sett på forskjeller mellom innvandrere og majoritetsbefolkning som har null års utdanning. Vi har da modellene:

$$lwage_i = \beta_1 + \beta_2 migrant_first + \beta_3 c_q09_c + \beta_4 c_q09_c_sq + \beta_5 yrqual + \beta_6 numscore1 + \beta_7 litscore1 + \beta_8 pslcore1 + \gamma_1 yrqual_mig + \epsilon_i \quad (4.1)$$

$$lwage_i = \beta_1 + \beta_2 migrant_first + \beta_3 c_q09_c + \beta_4 c_q09_c_sq + \beta_5 yrqual + \beta_6 numscore1 + \beta_7 litscore1 + \beta_8 pslcore1 + \gamma_2 num_mig + \gamma_3 lit_mig + \gamma_4 psl_mig + \epsilon_i \quad (4.2)$$

Da får vi disse verdiene i tabell 3: I 4.1 vil tolkningen være: Når dummyen for

innvandring settes lik null vil vi stå igjen med majoritetsbefolkningen. Da vil β_2 angi lønnsforskjellen mellom innvandrere og majoritetsbefolkningen, β_5 vil angi lønnsavkastning på utdanning for majoritetsbefolkningen. γ_1 er lønnsavkastningen på utdanning for innvandrere. Vi ser at for majoritetsbefolkningen er ett års ekstra utdanning forbundet med 4% økning i lønn, mens for innvandrere er ett års ekstra utdanning forbundet med 0.5% reduksjon i lønn. Dog er TS for $yr\text{squal}_{mig} = -0.8$ dette er for lavt til å forkaste en nullhypotese om at lønnsavkastning er forskjellig for innvandrere og majoritetsbefolkningen.

Tilsvarende i 4.2 når dummyen settes lik null vil β_2 angi lønnsforskjellen mellom innvandrere og majoritetsbefolkningen, β_6 vil angi lønnsavkastning på tallferdigheter for majoritetsbefolkningen, β_7 vil angi lønnsavkastning på leseferdigheter for majoritetsbefolkningen, β_8 vil angi lønnsavkastning på problemløsing for majoritetsbefolkningen. γ_2 er lønnsavkastningen på tallferdigheter for innvandrere, γ_3 er lønnsavkastningen på leseferdigheter for innvandrere, γ_4 er lønnsavkastningen på problemløsing for innvandrere. Vha. F-test kan vi teste en hypotese om at lønnsavkastningen for innvandrere på ferdigheter er forskjellig fra majoritetsbefolkningen. Fra stata:

```
. test num_mig lit_mig psl_mig

( 1) num_mig = 0
( 2) lit_mig = 0
( 3) psl_mig = 0

F( 3, 2661) = 0.89
Prob > F = 0.4475
```

Dette er lavere enn kritisk verdi. Vi kan ikke forkaste hypotese om at forskjeller i ferdigheter kan forklare lønnsgapet.

4.5 Begrensninger

I denne analysen er det ikke tatt hensyn til at innvandrere er systematisk ansatt i forskjellige sektorer. Det er naturlig å tenke seg at dette kunne stått for en del av lønnsgapet. Det er heller ikke gjort noen analyse av antall arbeidsledige, og ikke basert på kjønn. Det er naturlig å tenke at de menneskene som kommer fra mer kulturkonservative land enn Norge at familie er mer i fokus og at en forelder velger

å være hjemmeværende.

Vi har heller ikke kunnet sett på alder. Vi vet at når man er født har noe å si for hvor stor lønnsvekst man kan forvente å ha.

Tabell 3: Resultater av regresjonanalyse med interaksjon

VARIABLES	(1) F lwage	(2) G lwage
migrant_first	-0.04358 (0.01891)	-0.04903 (0.01861)
c_q09_c	0.01929 (0.00202)	0.01932 (0.00203)
c_q09_c_sq	-0.00026 (0.00004)	-0.00026 (0.00004)
yrsqual	0.04059 (0.00281)	0.03976 (0.00265)
numscore1	0.08248 (0.01305)	0.08890 (0.01387)
litscore1	-0.04266 (0.01346)	-0.05029 (0.01432)
pslscore1	0.04887 (0.00983)	0.05003 (0.01059)
num_mig		-0.05457 (0.03972)
lit_mig		0.06520 (0.04166)
psl_mig		-0.01103 (0.02782)
yrsqual_mig	-0.00552 (0.00693)	
Constant	2.32659 (0.04844)	2.33960 (0.04600)
Observations	2,672	2,672
R-squared	0.240	0.241

Standard errors in parentheses □

5 Oppsummering

Det er tydelig at det finnes et lønnsgap mellom innvandrere og majoritetsbefolkningen. Samtidig er det mange faktorer som reduserer gapet. Spesielt tallforståelse virker å ha en stor innvirkning på lønnsgapet, og når vi kontrollerer for dette reduseres gapet betydelig. Lønnsgapet varierte fra 17% til i overkant av 4.5%. Det er dog ikke funnet tegn på diskriminering i denne analysen basert på at innvandrere og majoritetsbefolkningen har samme avkastning på utdanning og ferdigheter. Basert på at vi ikke fant forskjell i avkastning på ferdigheter og utdanning, samt at lønnsgapet ble sterkt redusert når vi kontrollerte for noen faktorer, er det naturlig å tenke at det finnes ytterligere faktorer som kan redusere lønnsgapet ytterligere.

6 referanser

<https://www.ssb.no/arbeid-og-lonn/artikler-og-publikasjoner/lavere-lonn-blant-innvandrere>

<https://www.ssb.no/befolkning/artikler-og-publikasjoner/innvandrere-i-norge-sverige-og-danmark>

<https://www.ifau.se/sv/Press/Meddelanden/Invandrade-med-svensk-utbildning-har-lagre-lon-an-infodda/>

<https://www.skatteetaten.no/bedrift-og-organisasjon/arbeidsgiver/a-meldingen/om-a-ordningen/om-a-ordningen/>

R.L Thomas (2005): *Using Statistics in Economics*

