

Herman Lampe  
Kristoffer Fangberget Imset  
Magnus Wang Wold  
Nils Hadle Hidle

## **SØK2013 Bacheloroppgave i samfunnsøkonomi**

Lønnpåvirkningen av innvandring i Norge

Bacheloroppgave i Samfunnsøkonomi

Veileder: Costanza Biavaschi

Mai 2021



Herman Lampe  
Kristoffer Fangberget Imset  
Magnus Wang Wold  
Nils Hadle Hidle

## **SØK2013 Bacheloroppgave i samfunnsøkonomi**

Lønnsåvirkningen av innvandring i Norge

Bacheloroppgave i Samfunnsøkonomi  
Veileder: Costanza Biavaschi  
Mai 2021

Norges teknisk-naturvitenskapelige universitet  
Fakultet for økonomi  
Institutt for samfunnsøkonomi



**NTNU**

Kunnskap for en bedre verden



- *En stor takk til Costanza Biavaschi for god veiledning, raske tilbakemeldinger og verdifulle innspill.*

# Innholdsfortegnelse

1. Innledning .....	3
1.1 Kort om innvandringen i Norge frem til i dag .....	4
1.2 Arbeidsinnvandring - hvorfor Norge er et lukrativt mål .....	5
1.3 Problemstilling .....	7
2. Metode .....	7
2.1 Regresjonsanalyse .....	7
2.2 Hypotesetesting .....	9
3. Regresjon .....	10
3.1 Materiale.....	10
3.2 Styrker og svakheter i datasettet: .....	11
3.3 Regresjonsanalyse .....	12
3.4 Diskusjon.....	18
4. Oppsummering.....	20
5. Referanser .....	22

# 1. Innledning

Spørsmålet om hvorvidt innvandring påvirker lønnsnivået i et land er en kompleks, samfunnsøkonomisk debatt. Vi ser fra tidligere forskning at lønnsnivået innad i visse sektorer faller som følge av arbeidsinnvandring. I artikkelen til Bratsberg og Raaum (2013) avdekkes påvirkningen innvandringen har på lønnen innad i visse sektorer i Norge. Blant annet viser deres regresjonsanalyse at koeffisienten til snekkeryrket var på -0.8, som tilsier at en 1% økning i innvandring isolert sett gir snekkere en lønnsnedgang på 0.8%.

I analysen gjort av Borjas (2003) konkluderer han med at en 10 % økning i innvandring i USA reduserer lønnen til den innfødte befolkningen med 3-4%. På bakgrunn av analysen til Borjas (2003) har Ottaviano og Peri (2012) i sin analyse kommet frem til at innvandring gir en økning på 0.6% i lønnen til innfødte. Ottaviano og Peri (2012) kaster lys over effekten av substituerbarheten mellom innfødte og innvandrere med lignende utdanning og erfaring. Som følge av dette blir resultatet av innvandringens påvirkning på de innfødtes lønn positiv i motsetning til analysen til Borjas (2003).

Det er også gjort en analyse av dette i det norske arbeidsmarkedet. Røed (2005) peker i sin analyse på effekten innvandring har på lønnsnivået til den norskfødte delen av befolkningen. Modellen vektlegger hvordan forskjellige utdanningsnivåer mellom innvandrerne påvirker modellen, og splitter opp innvandrerne i to grupper: Utdannelse på mer enn 12 år, og mindre enn 13 år. Resultatene viste at innvandrergruppen med lav utdanning hadde en negativ effekt på det norskfødte lønnsnivået. På den andre siden hadde ikke-vestlige innvandrere med utdanning på 12 år eller mer en positiv effekt både på lønnsnivået til den norskfødte befolkningen og de vestlige innvandrerne. Satt opp mot hverandre, viste analysen at innvandrerandelen ikke påvirket lønnsdannelsen for en norsk arbeidstaker. Grunnen til dette kan ifølge Røed være at lønnseffekten blir tilslørt som følge av de motsatte sammenhengene gjeldende høy og lav utdanning.

Hvordan man tar for seg spørsmålet om hvorvidt innvandring påvirker lønnsnivået til de innfødte i et land har derfor stor innvirkning på hva regresjonsanalysen forteller oss. Vi har i vår analyse tatt utgangspunkt i innvandringens påvirkning på lønn segregert i de forskjellige kommunene i Norge i årene 2015-2019.

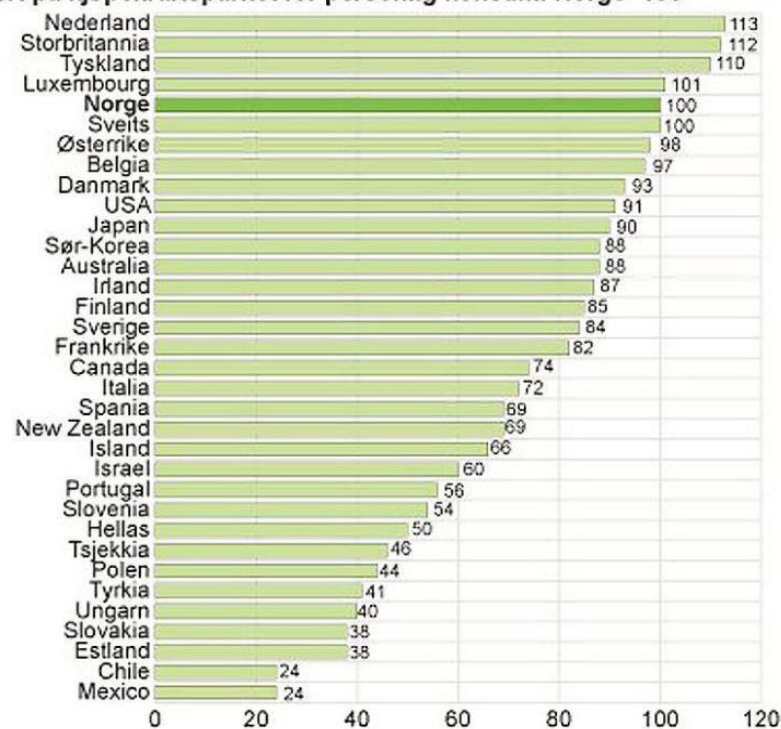
Vi benytter oss av en rekke spesifikasjoner i vår valgte modell. Dette inkluderer å kontrollere for økonomisk aktivitet, innbyggertall og dummyvariabler for kommuner og år. Vi gjorde dette for å mer effektivt isolere for innvandringens effekt på lønn. I løpet av analysen fant vi ingen konklusive bevis som støtter at innvandring har en effekt på lønnsnivået i Norge.

Bacheloroppgaven starter med en teoridel før vi kartlegger tidligere forskning og viser hva vi sammenlignet våre funn med. Deretter følger problemstillingen, oppbygningen og presentasjonen av modellen. Oppgaven avsluttes med en diskusjonsdel etterfulgt av en oppsummering.

## 1.1 Kort om innvandringen i Norge frem til i dag

Norge har i lang tid hatt innvandring fra våre nordiske og europeiske naboland, som svenske anleggsarbeidere og danske embetsmenn. Nettoinnvandringen i landet holdt seg relativt stabil fra 1967 til starten av 1980-tallet. Fra midten av 1980-tallet har innvandringen økt gradvis frem til 2012, hvor det ble nådd et maksimum på over 56 000 registrerte innvandringer med ikke-nordisk statsborgerskap. Etter 2012 har innvandringstallene gått noe ned. (Østby, 2017)

**Figur 2. Nivåindekser for bruttolønn ekskl. arbeidsgiveravgift, 2010. Basert på kjøpekraftsparitet for personlig konsum. Norge=100**

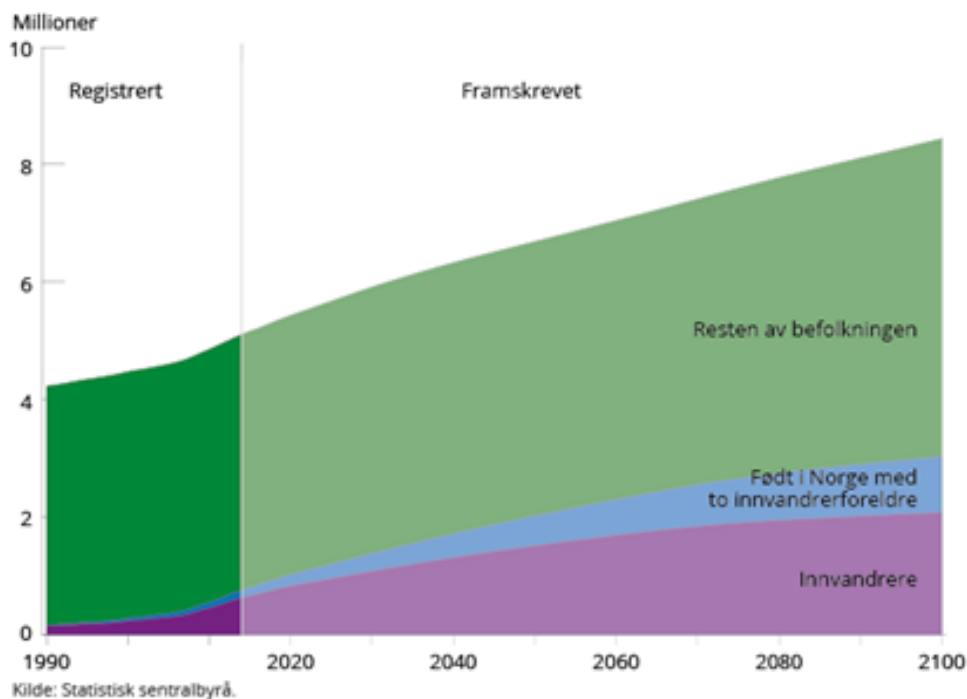


(Svennebye, 2011)



De siste tiårene har etterspørselen etter ufaglært arbeidskraft økt kraftig, spesielt i sektorer som jordbruk og anleggsvirksomhet. Hvis man ser på lønnsnivået i Polen ligger det på under halvparten av lønnsnivået i Norge, og i kombinasjon med EU-utvidelsen i 2004 hvor blant annet Polen og Litauen ble medlem av EU, har innvandring fra de ovennevnte landene økt kraftig.

Videre viser statistikken og antakelsene til SSB at innvandringen kan fortsette å øke i lang tid fremover.



(Tønnessen, 2016)

SSB antar at i 2060 kan 24% av befolkningen i Norge være innvandrere (Tønnessen, 2016). Hvorvidt innvandring har en påvirkning på norsk økonomi er derfor ikke et spørsmål som kun er relevant for dagens næringsliv, men også for den fremtidige økonomiske situasjonen i Norge.

## 1.2 Arbeidsinnvandring - hvorfor Norge er et lukrativt mål

Norge er et av de mest velstående landene i verden og har derfor et svært høyt lønnsnivå. Vi er også et av landene i verden med jevnest lønnsstruktur. Det vil derfor være de med relativt lav produktivitet som har størst insentiv til å komme til Norge. Empiri viser at jo større

sprang i velferden mellom Norge og innvandringslandene, desto større vil innvandringen være fra de respektive landene (Røed, 2005).

Grunnleggende økonomisk teori sier at flere arbeidstakere gir nedgang i lønn eller lavere lønnsvekst. Teorien er at større konkurranse rundt arbeidsplassene presser lønningene ned slik at flere kan ansettes (Bratsberg & Raum, 2013). Denne effekten er forskjellig blant ulike yrkesgrupper. Arbeidskraft som lett kan erstattes, slik som lavt utdannet arbeidskraft, er mest utsatt (Bratsberg & Raum, 2013). Problemet med å konkludere med at innvandring fører til lønnsvekst, er at arbeidsinnvandrere som regel søker seg til områder med stor vekst og etterspørsel etter arbeidskraft. I et slikt område er det naturlig med lønnsvekst, uavhengig av innvandring.

Studien til Bratsberg og Raum (2013) så blant annet på utviklingen fra 1998-2005. Her så man snekkeryrket som hadde mye arbeidsinnvandring, i motsetning til blant annet elektro og VVS. Som nevnt innledningsvis, hadde arbeidsinnvandring en negativ effekt på lønn i snekkeryrket, med en koeffisient på -0.8. Grunnen til at snekkeryrket opplevde så mye innvandring i forhold til elektro og VVS, er kravet om norsk utdanning for de sistnevnte. Vi ser altså støttende bevis på at innvandring kan påvirke lønn i utsatte, lavt utdannede yrker (Bratsberg & Raaum, 2013).

Tidligere forskning forteller oss hvorvidt innvandring påvirker lønnsnivået avhengig av hvor lett arbeidsgruppene kan erstatte hverandre (Ottaviano & Peri, 2012). Et spesifikt eksempel vi kan vise til er servicebransjen i Oslo. Her har innvandring fra Sverige stor innvirkning, og påvirker for det meste unge voksne. Bratsberg og Raum viser til en koeffisient på -0.88, altså vil en 1% økning i svensk innvandring gi 0.88% færre ansatte norske unge voksne (Bratsberg & Raum, 2013).

For å oppsummere viser tidligere forskning at innvandring har en påviselig effekt på lønnsnivået innad i sektorene. Det må spesifiseres at de store delene av innvandringen Norge har, altså fra Norden og Øst-Europa, i hovedsak påvirker lavt utdannede yrker. Disse lavt utdannede yrkene ser lavere lønnsvekst og større kamp om arbeidsplassene, noe de norske lavt utdannede arbeidstakerne ikke tjener på. Befolkningen ellers kan være tjent med innvandring, siden det er med på å senke prisvekst. Eksempelvis vil lavere lønninger i serviceyrkene, som nevnt ovenfor, føre til billigere tilbud for den gjennomsnittlige nordmann (Bratsberg og Raum, 2013).

### 1.3 Problemstilling

Vi velger å basere hoveddelen av bacheloroppgaven på en empirisk, kvantitativ analyse av følgende problemstilling:

*Påvirker innvandring lønnsnivået generelt i Norge?*

Gitt problemstillingen legger vi til grunn en nullhypotese om at innvandring ikke påvirker lønnsnivået i Norge. Vår forventning før utførelse av analysen er at innvandring har en effekt på lønnsnivået. Vi baserer denne forventningen på resultatene av tidligere forskning og samfunnsøkonomisk intuisjon.

## 2. Metode

I dette kapitlet skal vi gå gjennom teorien som legges til grunn for utførelsen av regresjonsanalysen. Vi starter med å gå gjennom generell teori for regresjonsanalyse, deretter utledes teori om hypotesetesting.

### 2.1 Regresjonsanalyse

Regresjonsanalyse er en kvantitativ analyse av sammenhenger mellom en avhengig variabel og en eller flere uavhengige variabler. Regresjonsanalysen settes opp som en lineær funksjon, hvor den avhengige variabelen  $Y$  defineres av en eller flere uavhengige variabler  $X$ . Vi vil bruke Minste Kvadraters Metode (MKM) som et verktøy for å analysere regresjonsanalysen. MKM er en metode som brukes for å finne sammenhenger mellom observerte verdier. I vårt tilfelle vil denne metoden gi en indikasjon på sammenhengen mellom lønn og de uavhengige variablene i regresjonsanalysen. Det settes opp som en lineær funksjon på formen:

$$Y = a + m_1x_1 + m_2x_2 + \dots + m_kx_k + u$$

$Y$  er den avhengige variabelen (endogen),  $a$  er konstantleddet som skjærer  $y$ -aksen og  $x$  er de uavhengige (eksogene) variablene. Effekten av de uavhengige variablene  $x_i$  har på  $Y$ , er gitt av  $m_i$ .

For at regresjonsanalysen skal holde, må vi legge fem forutsetninger til grunn:

### **MLR.1 – Linearitet**

Modellen skal være lineær på formen:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k + u,$$

hvor  $\beta_0, \beta_1, \dots, \beta_k$  er de ukjente parameterne vi er interesserte i, og  $u$  fanger alt som ikke blir forklart av modellen. Modellen må være lineær for at vi skal kunne utføre en regresjonsanalyse. (Woolridge, 2012, s. 83)

### **MLR.2 – Tilfeldig utvalg**

Vi ønsker at utvalget i modellen skal være tilfeldig for å unngå unøyaktighet og partiske data. (Woolridge, 2012, s. 84)

### **MLR.3 – Ingen perfekt kollinearitet**

Det skal ikke være perfekt kollinearitet mellom  $x_i$  leddene i modellen, og ingen konstante uavhengige variabler.

$Var(x_i) \neq 0$  må gjelde for at modellen skal holde. Ingen uavhengig variabel kan være en kombinasjon av to andre uavhengige variabler. MLR godtar noe korrelasjon mellom variablene, men perfekt korrelasjon fører til at MKM-metoden vil feile. (Woolridge, 2012, s. 84)

### **MLR.4 – Zero conditional mean**

$E(u|x_1, x_2, \dots, x_k) = 0$ . Feilleddet ( $u$ ) er et mål på andre faktorer som ikke har blitt tatt med i modellen. For at modellen ikke skal bryte med denne forutsetningen må den forventede verdien til feilleddet for alle  $x_i$  være lik 0. (Woolridge, 2012, s. 86)

### **MLR.5 – Homoskedastisitet**

$Var(u|x_1, x_2, x_3, \dots, x_k) = \sigma^2$ . Vi vil at data skal være homoskedastisk så ikke variansen til feilleddet ( $u$ ) blir ulik for ulike  $x_i$ . (Woolridge, 2012, s. 93)

## 2.2 Hypotesetesting

Vi skal foreta en hypotesetest for å se om innvandring gir utslag på lønnsnivået i landet. I løpet av en hypotesetest fastsettes en nullhypotese ( $H_0$ ) og en alternativhypotese ( $H_A$ ).

Formålet er å se om nullhypotesen holder ved et gitt signifikansnivå, eller om man kan forkaste nullhypotesen. Signifikansnivået er en parameter vi setter før utførelse av modellen, som sier hvor statistisk signifikant et resultat må være for å forkaste nullhypotesen. Et større signifikansnivå gir en større sannsynlighet for at nullhypotesen kan forkastes. Normalt settes signifikansnivået til 5% (Woolridge, 2012, s. 124).

Ved en hypotesetest kan man enten gjennomføre en ensidig test eller en tosidig test. Dette kommer an på om formålet med hypotesen er å sjekke om resultatene øker eller synker i forhold til nullhypotesen (ensidig test), eller om man vil sjekke begge samtidig (tosidig test). Ved en ensidig test vil forkastningsområdet kun være på en side av utvalgsfordelingen, mens ved en tosidig test vil forkastningsområdet fordeles likt på begge sider.

Vår hypotesetest vil ta utgangspunkt i en tosidig test, hvor vi tester hvorvidt innvandring påvirker lønnsnivået på generell basis i Norge, med et signifikansnivå på 5%.

Vi setter opp hypotesen vår slik:

$$H_0: \beta_1 = 0$$

$$H_A: \beta_1 \neq 0$$

Type 1- og type 2 feil er et viktig konsept man må ta stilling til når man utfører en hypotesetest. En type 1 feil er en «false positive» der en sann nullhypotese blir forkastet på grunn av feil i testprosedyren. På den andre siden er en type 2 feil «false negative», at en falsk nullhypotese ikke blir forkastet på grunn av feil i testprosedyren.

For å avgjøre hvorvidt man kan forkaste en nullhypotese eller ikke, må vi ta stilling til hvilke p- og t-verdier vi får fra datasettet. P-verdien sier hvor sannsynlig det er at vi begår en type 1 feil.

### 3. Regresjon

I dette kapittelet skal vi gjennomføre analysen av modellen, med utgangspunkt i teorien vi har lagt til grunn. Vi vil først utlede oppbygningen av datasettet, deretter diskuterer vi styrker og svakheter ved modellen. Avslutningsvis utføres en trinnvis regresjon der vi gradvis legger til flere variabler for å korrigere for feil, og diskuterer resultatene.

#### 3.1 Materiale

Data for datasettet er hentet fra SSB, omformet til riktig format i Excel og satt sammen i STATA. Vi har valgt å se på kommunale data med kommunefordelingen i årene 2018-2019. Dette fordi kommunesammenslåingen medførte endringer i kommuneutformingene. Videre utleder vi hvordan vi bygget opp datasettet og hvilken teori som ble lagt til grunn for analysen.

##### **Lønn:**

Hentet fra datasett 12852 fra SSB (månedslønn, 2015-2019, gjennomsnitt, arbeider i regionen, alle kommuner 2018-2019). Gir oss gjennomsnittlig månedslønn for arbeidstakere på kommunebasis.

##### **Bosatte:**

Hentet fra datasett 07459 fra SSB (2015-2019, alle kommuner 2018-2019). Her hentes innbyggertall for hver respektive kommune.

##### **Innvandring:**

Hentet fra datasett 07110 fra SSB (2015-2019, alle kommuner 2018-2019). Henter ut antall innvandrere i hver kommune.

##### **Sammensetting av datasettet:**

Setter alle datasett på formen (kommuner, årstall og data) i Excel. Overfører deretter til STATA og konverterer til «.dta» filer.

Vi bruker STATA til å slå sammen datasettene med basis i kommuner og årstall for å få all data inn i samme fil. Denne datafilen brukes deretter til å utføre regresjonsanalysen.

## VARIABLER

Navn	Forklaring	Type
<b>loglønn</b>	Naturlig logaritme av månedslønn	Kontinuerlig
<b>logbosatte</b>	Naturlig logaritme av antall bosatte i en kommune	Kontinuerlig
<b>innvandring</b>	Innvandring fordelt på kommunene i landet	Kontinuerlig
<b>seksten/syttten/atten/nitten</b>	Årsvariabler, 2016-2019	Dummy
<b>kommune</b>	Kommuner i landet	Kategori

*Tabellen viser en oversikt over variabler som er brukt i den endelige modellen.*

### 3.2 Styrker og svakheter i datasettet:

#### **Styrker:**

Datasettet består av 1955 observasjoner. Gitt antall observasjoner og kilden data er hentet fra (Statistisk Sentralbyrå), er det rimelig å anta at data er pålitelig. Observasjonene presenterer lønnen til befolkningen fordelt mellom kommunene over en tidsperiode på fem år (2015-2019).

Datasettet er longitudinelt, som kan føre til at vi klarer å korrigere for særegenheter i data for enkelte årstall og/eller kommuner.

#### **Svakheter:**

Selv om de viktigste variablene som representerer innvandring og lønn er med i datasettet, er det usannsynlig at modellen har klart å fange opp alle faktorer som vil påvirke denne

sammenhengen. Nasjonale faktorer som politikk og demografi, samt globale faktorer som krig, global markedsaktivitet og oljepris er variabler som kan påvirke innvandring og lønn, men som kan være vanskelige å ta med i beregningen.

Et annet sentralt problem er at datasettet mangler noen observasjoner for enkelte kommuner. Trolig har kommunesammenslåingene som skjedde i perioden 2015-2019 ført til noen av inkonsekvensene i datasettet. Det kan dermed være en skjevhet i utelatelsen som modellen fanger. Det er også nødvendig å merke seg at data for bosatte i Norge har en stor skjevhet som kan føre til problemer i regresjonsanalysen. Dette blir diskutert videre i kapittel 3.4.

### 3.3 Regresjonsanalyse

I dette delkapittelet utleder vi hovedresultatene av analysen vår, resultatet av regresjonsanalysen er presentert i tabell 1 under:

Tabell 1:

VARIABLES	(1) lønn	(2) lønn	(3) lønn	(4) lønn	(5) loglønn
innvandring	0.104 (0.00741)	-0.310 (0.0324)	-0.323 (0.0258)	0.467 (0.218)	-9.86e-07 (1.15e-06)
bosatte		0.0962 (0.00734)	0.0989 (0.00584)	-0.157 (0.0872)	
seksten			1,295 (163.9)	1,269 (37.96)	0.0338 (0.000934)
sytten			2,469 (163.6)	2,399 (38.86)	0.0627 (0.000950)
atten			3,395 (158.8)	3,418 (39.49)	0.0877 (0.000945)
nitten			4,925 (158.7)	4,910 (40.78)	0.123 (0.000950)
logbosatte					-0.0433 (0.0229)
Constant	40,368 (67.39)	39,891 (74.21)	37,370 (120.1)	39,210 (730.1)	10.92 (0.195)
Kommune dummies	No	No	No	Yes	Yes
Observations	1,955	1,955	1,955	1,955	1,955
R-squared	0.092	0.165	0.474	0.978	0.977

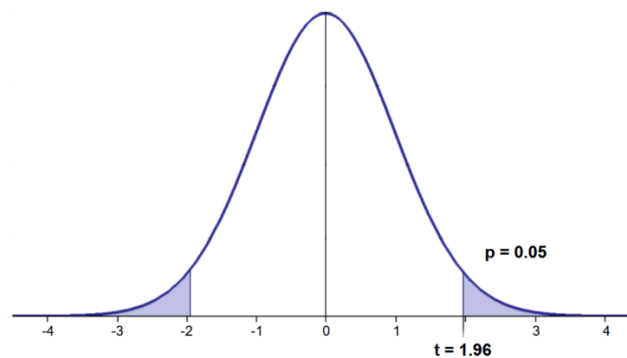
*Tabellen viser koeffisientene mellom de endogene og eksogene variablene. Standardfeil befinner seg i parentesene.*



Basert på datasettene fra SSB kan vi starte med å sette opp en enkel regresjon mellom lønn og total innvandring til kommunene.

$$Lønn = \beta_0 + \beta_1 Innvandring + u$$

Variabelen *innvandring* måler hvor mange innvandrere som kommer til hver kommune hvert år. Ved hjelp av denne variabelen kan vi se om det er samvariasjon mellom økning i innvandrertall og snittlønn i en kommune. Siden vi bruker et 5% signifikansnivå, får vi en t-distribusjon med frihetsgrad lik  $1955 - 429 - 1 = 1525$  ( $df = n - k - 1$ ).



Vi gjennomfører en tosidig test med 2.5% signifikansnivå på hver side. Med andre ord så tester vi imot en t-score på 1.96 og forkaster nullhypotesen når:

$$|TS| > 1.96$$

Hvis man starter med kun disse to variablene får vi en regresjon som gir innvandring en koeffisient på 0.104 og en standard feil på 0.00741 som vises i tabell 1 (1) øverst i kapitlet.

Regner ut teststatistikk:

$$TS = \frac{0.104}{0.00741} = 14$$

Siden  $TS > 1.96$  anser vi variablene som statistisk signifikante, men det er mange variabler som korrelerer både med lønn og innvandring samtidig, vi trenger derfor å kontrollere for disse for å unngå brudd på MLR.4.

### **Totale bosatte**

En av de letteste faktorene å korrelere med lønn og innvandring samtidig vil være bosatte. Vi vet fra MLR.4 at forutsetningen blir brutt hvis man har variabler i  $u_i$  som forklarer både den

avhengige og en eller flere av de uavhengige variablene. Bosatte er en faktor som trolig skaper dette problemet. Antallet som bor i en kommune påvirker trolig både lønn og innvandring i seg selv, den må dermed inkluderes og vi får modellen:

$$Lønn = \beta_0 + \beta_1 innvandring + \beta_2 bosatte + u.$$

Vi ser i Tabell 1 (2) koeffisientene til regresjonen som også inneholder bosatte. Vi får her  $\hat{\beta}_1 = -0.310$ ,  $\hat{\beta}_2 = 0.0962$ . TS for disse blir da:

$$TS_1 = -\frac{0.310}{0.324} = -9.56 \rightarrow |-9.56| > 1.96$$

$$TS_2 = \frac{0.0962}{0.007342} = 13.10 > 1.96$$

Disse verdiene er også signifikante og nullhypotesen ville her blitt forkastet. I motsetning til forrige regresjon blir innvandringskoeffisienten negativ mens befolkningskoeffisienten blir positiv. Vi får at hvis én ekstra innvandrer kommer til kommunen, forventes månedslønnen å gå ned med 9.56 kroner. Derimot ser vi at 1 ekstra beboer medfører en lønnsøkning på 13.10 kr per måned. Disse to sammenlagt medfører at en innvandrer bør forventes å øke snittlønnen med  $13.10 - 9.56 = 3.54$ kr.

## Årstall

Siden vi analyserer et longitudinelt datasett må det også tas hensyn til faktorer som påvirkes av tid. Økonomisk aktivitet og inflasjon er to faktorer som teoretisk sett har innvirkning på nominell lønn. BNP og KPI blir dermed variabler som kan forklare lønnsendringer over tid. Eksempelvis vil positiv inflasjon medføre en forventet oppgang i lønn siden arbeidstakere som regel forhandler seg vekk fra reallønnsnedgang. Siden KPI og BNP i stor grad måles på nasjonalt nivå vil disse variablene ha stor korrelasjon med årstallsdummyvariabler. Dermed bruker vi heller de tidligere nevnte dummyvariablene. Disse dummyvariablene vil også fange andre faktorer som endrer seg på årlig basis. Vi får da en regresjon på formen:

$$lønn = \beta_0 + \beta_1 innvandring + \beta_2 bosatte + \beta_3 seksten + \beta_4 sytten + \beta_5 atten + \beta_6 nitten + u$$

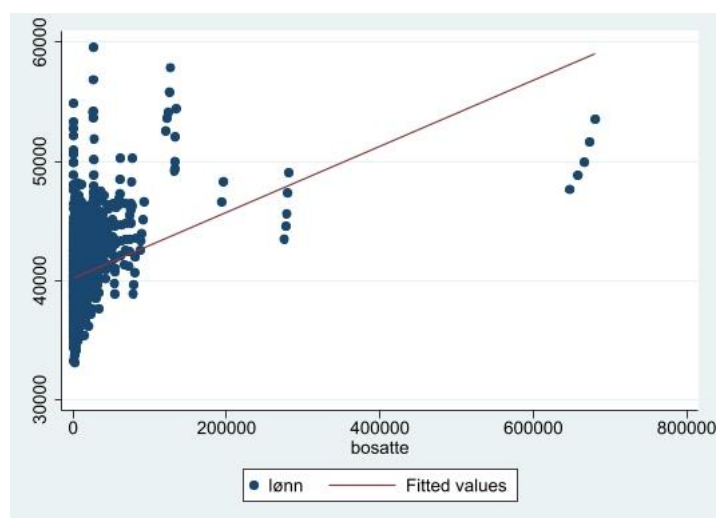
Regresjonen vises i tabell 1 (3).  $\hat{\beta}_3$  til og med  $\hat{\beta}_6$  viser da lønnsvekst som skyldes årlige endringer. Her blir også henholdsvis t-scorene til  $\hat{\beta}_1$  og  $\hat{\beta}_2$  større.

### Kommuner

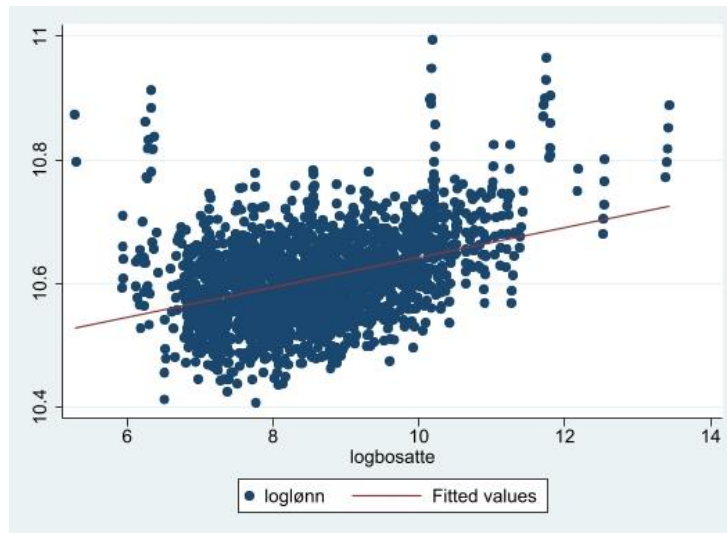
Innad i kommunene i Norge er det også mange forskjellige faktorer som kan korrelere med lønn og innvandring samtidig. Eksempelvis kan en kommunes geografiske forhold, eller tilgang til spesielle ressurser, korrelere med både innvandring og lønn. Derfor må vi også kontrollere for forskjeller mellom kommunene. Vi oppretter dummyvariabler, en for hver kommune. Fra tabell 1 (4) ser vi at innvandringskoeffisienten blir positiv, med en verdi på 0.467 og en t-verdi på 2.14. Bosattekoeffisienten blir nå negativ med en verdi på -0.15 og t-verdi på -1.80. Vi ser her at de ovennevnte koeffisientene ikke er signifikante for et 5% signifikansnivå.

### Naturlig logaritme

Siden Norge har store variasjoner i befolkningstall mellom kommuner, ønsker vi å korrigere for den naturlige heteroskedasiteten i datasettet for å ikke bryte med MLR.4 og MLR.5. En av hovedproblemene er at storbyene har relativt høye innbyggertall, hvilket medfører ujevnheter i data.



En av hovedmetodene for å fikse denne skjevheten er å ta den naturlige logaritmen av variablene. Ved å ta logaritmen blir data mer komprimert og vi får en jevnere fordeling av variansene for  $u_i$  for forskjellige  $x$ .



Regresjonen med loglønn og logbosatte vises i tabell 1 (5). Vi ser da at innvandring får en koeffisient på  $-9.86e-07$ . Siden innvandring er en kontinuerlig variabel, må den transformeres med formelen  $100 * (\exp(\beta_j \Delta x_j) - 1)$  for å leses mot log. Vi setter inn våre tall og får:

$$100 * (e^{-9.86e-07} - 1) \approx -0.00009$$

Med andre ord vil én ekstra innvandrer i gjennomsnitt forvente å redusere snittlønn med 0.00009%. T-verdien til innvandring blir:

$$TS_1 = \frac{-9.86e-07}{1.15e-06} = -0.86$$

Siden  $|TS_1| < 1.96$  kan vi konkludere med at innvandring ikke er statistisk signifikant i modellen. P-verdien tilsvarer et signifikansnivå på 39% som vår analyse ikke godtar.

Logbosatte har en koeffisient på  $-0.0433$ . Ettersom loglønn og logbosatte begge er logaritmiske variabler, trenger vi ikke å gjennomføre noen transformasjoner for å se påvirkningen. Vi finner at dersom bosatte endres med 1%, forventes også snittlønn å endre seg med  $-0.0433\%$ . Videre regner vi ut t-verdien til logbosatte og får:

$$TS_B = -\frac{0.0433}{0.0229} = -1.89 \rightarrow |-1.89| < 1.96$$

Teststatistikken innebærer med andre ord at også logbosatte er en statistisk insignifikant variabel.

Årstallsdummyvariablene må også transformeres for å tolkes i forhold til logaritmeformen til lønn. Med 2015 som basisår forteller disse variablene oss forventet lønnsvekst fra 2015 basert på året man er i, alt annet likt.

Hvis vi ønsker å finne virkningen 2016-variabelen har på lønn, kan den regnes ut slik:

$$100(\exp(0.0337692) - 1) \approx 3.4\%$$

Vi ser at den isolerte effekten på månedslønn fra 2015 til 2016 er 3.4%. Med andre ord forventer vi at gjennomsnittlig månedslønn i 2016 er 3.4% høyere enn i 2015 i snitt, ceteris paribus. Alle årstallsdummyvariabler har signifikante t-verdier.

Siden variablene logbosatte og innvandring begge har en høy t-verdi kan vi gjennomføre en f-test for å sjekke om de sammen er signifikante. Formulerer dermed en nullhypotese for F-testen:

$$H_0: \beta_1 = \beta_2 = 0$$

$$H_1: \text{ikke } H_0$$

Utfører testen i STATA og får resultatet:

```
( 1) logbosatte = 0
( 2) innvandring = 0

F( 2, 1526) = 2.91
Prob > F = 0.0550
```

Vi ser fra testen at det er en 5.5 prosent sannsynlighet for å gjøre en type 1 feil ved å forkaste nullhypotesen. Med andre ord forkaster vi ikke nullhypotesen ved et 5 prosent signifikansnivå, og kan dermed ikke vedta at disse variablene sammen påvirker lønn.

Vi vil også kommentere korrelasjonen mellom de forskjellige variablene. En viktig korrelasjon å se på er den mellom bosatte og innvandring da disse er nært beslektet. Variablene har en korrelasjonskoeffisient på 0.98, noe som kan være bekymringsverdig fordi kollinearitet er noe vi vil ha så lite av som mulig. Om vi derimot ser på logbosatte og innvandring er denne betydelig mindre, og det er disse variablene vi bruker i modellen.

Tabell 2:

	loglønn	lønn innvandring	logbosatte	bosatte	femten	seksten	sytten	atten	nitten	
loglønn	1.0000									
lønn	0.9976	1.0000								
innvandring	0.2883	0.3035	1.0000							
logbosatte	0.3915	0.3922	0.4201	1.0000						
bosatte	0.3407	0.3555	0.9757	0.5423	1.0000					
femten	-0.4153	-0.3933	-0.0053	0.0015	-0.0023	1.0000				
seksten	-0.1941	-0.1911	-0.0015	0.0030	-0.0009	-0.2334	1.0000			
sytten	-0.0005	-0.0066	0.0026	0.0040	0.0016	-0.2342	-0.2342	1.0000		
atten	0.1612	0.1511	0.0009	-0.0019	0.0004	-0.2531	-0.2531	-0.2540	1.0000	
nitten	0.4196	0.4117	0.0032	-0.0062	0.0011	-0.2535	-0.2535	-0.2543	-0.2749	1.0000

Tabellen viser en matrise som inneholder korrelasjonskoeffisientene mellom variablene.

### 3.4 Diskusjon

I en regresjonsmodell er det viktig å diskutere de forskjellige forutsetningene vi har, og hvordan de blir oppfylt i modellen.

Den mest grunnleggende forutsetningen er at modellen er lineær:

$$\text{Log}(\text{Lønn}) = \beta_0 + \beta_1 \text{Innvandring} + \beta_2 \text{Bosatte} + \beta_3 2016 + \beta_4 2017 + \beta_5 2018 + \beta_6 2019 + \beta_7 \text{Kommuner} + u,$$

*Kommuner representerer her et fullt sett med dummyvariabler, en for hver kommune.*

Tilfeldig utvalg er relevant i modellen vår siden det er noen ting vi ikke kan kontrollere for. All data i datasettet er hentet fra SSB og det er flere tilfeller av manglende data for diverse kommuner i datasettene vi brukte på lønn, innvandring og bosatte. Siden det er visse kommuner vi ikke kunne få med i modellen, vet vi ikke om datasettet bryter med denne forutsetningen. Til tross for dette har vi inkludert alle data vi fant i årsperioden 2015-2019 for alle kommunene som hadde observasjoner. Som tidligere diskutert er trolig kommunesammenslåingene i perioden 2015-2019 noe av grunnen til denne inkonsekvensen.

Ingen av variablene våre er konstante, og ingen av variablene har perfekte lineære forhold. Det er kollinearitet i datasettet, spesielt mellom antall bosatte i kommunene og antall innvandrere til kommunen. Spesielt med tanke på at om innvandring i en kommune øker med 1 vil også bosatte øke med 1. Dette 1 til 1 forholdet er ikke like signifikant den andre veien,

og det er også andre faktorer som påvirker bosatte slik som migrering. Med dette tatt i betraktning kan vi konkludere med at det ikke er noe perfekt kollinearitet i datasettet.

Den viktigste betingelsen i en regresjonsanalyse er at det ikke er kovarians mellom «u» og de uavhengige variablene. Denne betingelsen er sjelden oppfylt i modeller som tar for seg data fra den virkelige verden. I de aller fleste tilfeller er denne betingelsen brutt. Den største faktoren som kan føre til dette er “omitted variable bias”. Dette forekommer, som tidligere nevnt, når en viktig variabel som har korrelasjon både med en eller flere uavhengige variabler og den avhengige variabelen er utelatt fra modellen (Woolridge, 2012, s. 86). Som tidligere diskutert kan politikk være en slik variabel i vårt tilfelle. Politikk som er spesifikk for enkelte kommuner kan bli fanget opp av kommunedummyene i modellen vår, dette fordi politikk kan variere mellom kommuner i datasettet. Siden vi kun ser på fem år, er det vanskelig å fange opp årlige endringer i den landsdekkende politikken. Dette fordi det kan være et for lite tidsperspektiv til å fange opp nasjonale politiske endringer. Hvis man for eksempel ser på innvandrings - og økonomipolitikken til FrP kontra Rødt, vil dette være en faktor med direkte påvirkning på både lønn og innvandring. Hvilke partier som sitter i regjering har av den grunn stor påvirkningskraft for hva slags innvandringspolitikk som gjelder i landet. Dette er en faktor som kan forårsake et brudd på MLR.4 som er utelatt fra modellen grunnet kompleksiteten rundt inklusjonen av en slik variabel.

En variabel som var tiltenkt å være med i regresjonen, men som ble utelatt, er arbeidsledighet. Arbeidsledighet er teoretisk sett korrelert til både lønnsnivået og innvandring, og ville derfor vært en logisk variabel å ha med i modellen. Arbeidsledighet ble utelatt fra modellen fordi det er en endogen variabel, som betyr at den blir forklart av andre, eksogene variabler. Å bruke arbeidsledighet til å forklare endringer innvandring har på lønn ville dermed gjort modellen partisk, og vært et brudd på MLR.4.

Et annet mulig problem med datasettet er at vi ser på månedslønn istedenfor timelønn. Dette kan gjøre at vi mister faktorer innad i lønnsbildet. Den viktigste faktoren her er trolig arbeidstimer. Det vil være vanskelig å vite hvorvidt man ser på lønn eller arbeidstid om denne faktoren ikke fanges i modellen. Innvandring kan føre til større arbeidstilbud som igjen gjør at arbeiderne jobber mindre uten at timelønnen nødvendigvis endrer seg. Innvandring kan også ha den effekten at konsumentgruppen øker, som igjen fører til at bedriftene må produsere mer, derav en økning i arbeidstimer. Hvorvidt innvandring påvirker timelønnen er dermed uklart fordi vår analyse kun ser på månedslønn.

Den siste betingelsen vi vil diskutere er betingelsen om homoskedastisitet. Det vil si at variansen til feilleddet i modellen skal være konstant, gitt vilkårlige verdier for de uavhengige variablene (Woolridge, 2012, s. 93). Som tidligere diskutert prøvde vi å minimere hetroskedastisiteten i datasettet ved å gjøre en log funksjon på lønn og bosatte, men det er ikke sikkert at dette er tilstrekkelig for å kontrollere for hetroskedastisiteten i datasettet.

Tidligere forskning viser som tidligere nevnt til motstridende resultater, hvor Borjas (2003) fant en negativ effekt på lønnsnivået som følge av innvandring, mens Ottaviano og Peri (2012) ni år senere fant en positiv effekt på lønnsnivået. Grunnen til at tidligere forskning viser til motstridende resultater, samt at vi i vår regresjonsanalyse ikke kom frem til et konklusivt svar, skyldes trolig kompleksiteten til problemstillingen. Vi fant i vår regresjonsanalyse flere faktorer som antakeligvis har innflytelse på modellen, som var for omfattende å inkludere i datasettet. Som et resultat av dette, kan modellen virke noe unyansert da forutsetningene for regresjonsanalysen virker vanskelig å overholde.

## 4. Oppsummering

Regresjonsanalysen hadde som mål å finne en sammenheng mellom lønnsnivå og innvandring. Analysen tok høyde for flere faktorer som kunne påvirke lønnsnivå for å finne den isolerte påvirkningskraften innvandring har på lønn.

Sluttresultatet vises i tabell 1 (5), og det er her vi fant effekten innvandring har på lønn. Vi fant at koeffisienten til variabelen «*Innvandring*» gitt lønn var  $-9.86e-07$ . Koeffisienten viser hvor mye lønn endres når total innvandring øker med 1, hvilket betyr at dersom antall innvandrere øker med 1 gir dette en  $-0.0000986\%$  endring i månedslønn.

$$100(\exp(-9.86e - 07) - 1) \approx -0.0000986$$

Selv om koeffisienten viste at *Innvandring* gir en liten endring i lønnsnivået, er ikke dette nødvendigvis en sann verdi. For å avgjøre hvorvidt koeffisienten er signifikant, må vi se på p- og t-verdien til innvandring.

I modellen finner vi at *Innvandring* har en p-verdi på 0.39 og en t-verdi på -0,86. Dette tilsier at det er 39% sjans for at vi begår en type 1 feil. I vår hypotesetest satt vi signifikansnivået til 5%, som er lavere enn p-verdien vi finner i vår regresjonsanalyse.



Vi kan derfor konkludere med at vi ikke finner noen beviselige endringer i lønn gitt *Innvandring*, og kan ikke forkaste nullhypotesen.

## 5. Referanser

- Borjas, G. (2003, November 1). *Academic.oup.com*. Hentet April 12, 2021 fra Oxford Academic: <https://academic.oup.com/qje/article/118/4/1335/1925108?login=true>
- Bratsberg, B., & Raaum, O. (2013). *Frisch UiO*. Hentet April 16, 2021 fra Frisch UiO: [https://www.frisch.uio.no/publikasjoner/pdf/Samfunnsokonomien\\_Bratsberg\\_Raaum\\_3\\_2013.pdf](https://www.frisch.uio.no/publikasjoner/pdf/Samfunnsokonomien_Bratsberg_Raaum_3_2013.pdf)
- Ottaviano, G., & Peri, G. (2012, Februar 1). *academic.oup.com*. Hentet April 16, 2021 fra Oxford Academic: <https://academic.oup.com/jeea/article/10/1/152/2182016?login=true>
- Røed, M. (2005, April). *samfunnsforskning.brage.unit.no*. Hentet April 20, 2021 fra Samfunnsforskning.no: [https://samfunnsforskning.brage.unit.no/samfunnsforskning-xmlui/bitstream/handle/11250/177469/R\\_2005\\_4.pdf?sequence=3&isAllowed=y](https://samfunnsforskning.brage.unit.no/samfunnsforskning-xmlui/bitstream/handle/11250/177469/R_2005_4.pdf?sequence=3&isAllowed=y)
- Statistisk Sentralbyrå. (2021). 07110: Innvandrere, etter landbakgrunn (verdensdel) og kjønn (K) 1970 - 2021. Hentet April 1, 2021 fra <https://www.ssb.no/statbank/table/07110>
- Statistisk Sentralbyrå. (2021, Februar 23). 07459: Alders- og kjønnsfordeling i kommuner, fylker og hele landets befolkning (K) 1986 - 2021. Norge. Hentet April 1, 2021 fra <https://www.ssb.no/statbank/table/07459/>
- Statistisk Sentralbyrå. (2021). 12852: Kommunefordelt månedslønn, etter bosted, arbeidssted, alder og kjønn (K) 2015 - 2020. Hentet April 1, 2021 fra <https://www.ssb.no/statbank/table/12852>
- Steinkellner, A. (2020, Mars 9). *SSB.no*. Hentet April 14, 2021 fra SSB: <https://www.ssb.no/befolkning/artikler-og-publikasjoner/nesten-15-prosent-er-innvandrere>
- Svennebye, L. (2011, September 12). Vanskelig å sammenlikne kjøpekraft. Norge. Hentet April 24, 2021 fra SSB: <https://www.ssb.no/priser-og-prisindekser/artikler-og-publikasjoner/vanskelig-aa-sammenlikne-kjopekraft>
- Tønnessen, M. (2016, Juni 21). *SSB.no*. Hentet April 20, 2021 fra SSB: <https://www.ssb.no/forskning/demografi-og-levekaar/befolkningsutvikling-flytting-og-dodelighet/hvor-mange-innvandrer-til-norge-framover>
- Woolridge, J. M. (2012). *Introductory Econometrics A Modern Approach 5th Edition*. Mason: Cengage Learning.
- Østby, L. (2017, September 18). Arbeidsinnvandrere – fra hovedstadsfenomen til vanlig syn i hele landet. Hentet April 30, 2021 fra SSB: <https://www.ssb.no/befolkning/artikler->

og publikasjoner/arbeidsinnvandrere fra hovedstadsfenomen til vanlig syn i hele landet

