

Grongstad, Rikke Urstad
Heggem, Martine
Sollie, Berit Elise

Lønnsforskjeller mellom urbane og rurale områder i Norge

En empirisk analyse av hvordan ulike demografiske faktorer påvirker lønnsforskjeller mellom urbane og rurale områder i Norge

Bacheloroppgave i Samfunnsøkonomi

Veileder: Colin Peter Green

Mai 2024

Grongstad, Rikke Urstad
Heggem, Martine
Sollie, Berit Elise

Lønnsforskjeller mellom urbane og rurale områder i Norge

En empirisk analyse av hvordan ulike demografiske faktorer påvirker lønnsforskjeller mellom urbane og rurale områder i Norge

Bacheloroppgave i Samfunnsøkonomi
Veileder: Colin Peter Green
Mai 2024

Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
Institutt for samfunnsøkonomi



Kunnskap for en bedre verden

Sammendrag

Denne empiriske analysen er utført for å undersøke i hvilken grad de demografiske faktorene geografisk område, kjønn, alder, og utdanning kan ha en påvirkning på lønnsforskjellene mellom urbane og rurale områder i Norge. Landets kommuner per 2022 er definert som urbane eller rurale ut fra sentralitetsindeksen til Statistisk Sentralbyrå. Tverrsnittsdata benyttes da det ønskes å fokusere på nåtidens lønnsforskjeller i geografiske områder, og ikke forskjellene over tid. Det er utført en deskriptiv analyse for å enklere fremstille tallene fra datagrunnlaget. Metoden som benyttes for å utføre regresjonsanalysen er minste kvadraters metode, og oppgavens modeller består av to enkle lineære regresjonsmodeller og tre multiple. Resultatene av analysen fastslår at oppgavens valgte demografiske faktorer har en signifikant betydning i ulik grad for lønnsforskjellen mellom landets urbane og rurale områder. Uavhengig av hvilke faktorer som ytterligere inkluderes vil geografisk område være av signifikans for lønnsforskjellen. På bakgrunn av dette resultatet fremkommer det en generell forventning om at lønnen i urbane kommuner er høyere enn i rurale. Videre er utdanning den faktoren som har størst effekt på lønnsforskjellen, og det fastslås at menn har høyest nytte av å arbeide i urbane områder, da de oppnår større økning lønn i urbane områder enn kvinner. Alder har generelt lite betydning for lønnsforskjellene. Hypotese- og robusthetstestene av modellene konstaterer at resultatene kan anses som solide.

Abstract

This empirical analysis has been performed to examine the extent to which the demographic factors geographic area, gender, age, and education can influence wage differences between urban and rural areas in Norway. The country's municipalities as of 2022 are defined as urban or rural based on the centralization index made by Statistisk Sentralbyrå. Cross-sectional data is used to focus on the current wage differences in geographic areas, rather than differences over time. A descriptive analysis has been performed to present the dataset more clearly. The method used to conduct the regression analysis is the least squares method, and the models in the study consist of two simple linear regression models and three multiple ones. The results confirm that the chosen demographic factors significantly impact the wage gap between urban and rural areas. Regardless of which factors are further included, geographic area remains significant for the wage difference. Based on this result, there is a general expectation that wages in urban municipalities are higher than in rural ones. Education is the factor that has the highest effect on wage differences, and it is established that men have a higher benefit of working in urban areas, because they have a higher increase in pay in urban areas compared to women. Age generally has a small effect on the urban-rural wage gap. Hypothesis and robustness tests of the models confirm that the results from the analysis can be considered as solid.

Innholdsfortegnelse

1. Introduksjon	1
1.1 Problemstilling.....	1
1.2 Struktur	2
2. Teoretisk rammeverk	3
2.1 Tidligere litteratur	3
2.2 Oppgavens bidrag til litteraturen	5
3. Datamateriale	7
3.1 Datagrunnlag.....	7
3.2 Definisjon av urbane og rurale områder	8
3.3 Deskriptive analyser	9
3.4 Styrker og svakheter ved datagrunnlaget	12
4. Metode	14
4.1 Minste kvadraters metode (MKM).....	14
4.2 Enkel og multipl lineær regresjonsmodell (SLR og MLR).....	14
4.3 MLR-antakelser.....	15
4.4 Determinasjonskoeffisienten.....	17
4.5 Hypotesetesting	18
5. Regresjonsanalyser	20
5.1 Forklaring av variabler	20
5.2 Modellene.....	20
5.3 Valg av funksjonsform.....	21
5.4 Modell (1): Bylønnspremie	22

5.5 Modell (2) Gjennomsnittlig månedslønn med hensyn til utdanning og kjønn	24
5.6 Modell (3): Bylønnspremie med hensyn til alder.....	27
5.7 Regresjonsresultat.....	29
5.8 Robusthetsanalyse.....	30
6. Diskusjon	33
6.1 Tolkning av regresjonsresultater	33
6.2 Oppgavens anbefaling til videre forskning	35
6.3 Feilkilder.....	36
7. Konklusjon.....	37
8. Referanseliste.....	39
9. Vedlegg.....	41
Vedlegg 1: Forklaring av variabler og koeffisienter i regresjonsmodellene.....	41
Vedlegg 2: Sammenstilling av regresjonsresultater	47

1. Introduksjon

Lønnsforskjeller er et omdiskutert tema i dagens samfunn og det finnes mye litteratur om emnet. Denne empiriske analysen tar for seg de demografiske faktorene geografisk område, alder, kjønn og utdanning, og deres effekt på lønnsforskjellen mellom urbane og rurale områder i Norge. Områdene defineres ved bruk av sentralitetsindeksen, en indeks utviklet av Statistisk Sentralbyrå (SSB). Datagrunnlaget baseres på statistikk hentet fra SSB. Enkel og multipl lineær regresjonsmodell benyttes for å beregne de demografiske faktorene på lønn. Regresjonsresultatene hypotesetestes for å undersøke om de demografiske faktorene er av statistisk signifikans, og testes videre for robusthet. Dette danner grunnlag for å tolke hvorvidt regresjonsresultatene sammenfaller med tidligere litteratur, og for å kunne påpeke denne oppgavens bidrag til litteraturen.

1.1 Problemstilling

Som bachelorstudenter på vei ut i arbeidsmarkedet, anses det som svært aktuelt og interessant å undersøke hvilke demografiske faktorer som påvirker lønnsforskjeller i urbane og rurale områder. Oppgaven kan bidra til økt kunnskap om bylønnspremie, noe som vil være nyttig å tilegne seg ved avveining om arbeidssted.

Etter å ha undersøkt tidligere publikasjoner om det aktuelle temaet, kommer det frem ulike tilnærminger til lønnsforskjeller. Kjønn, alder, utdanning og geografisk område er eksempler på faktorer litteraturen undersøker om har påvirkning på lønn. Det kan være interessant å analysere lønnsforskjellene innad i Norge, da dette kan være et aktuelt og nyttig bidrag til allerede eksisterende empiri om lønnsforskjeller. På bakgrunn av disse argumentene vil denne oppgavens problemstilling være hvordan ulike demografiske faktorer påvirker lønnsforskjeller i urbane og rurale områder i Norge.

1.2 Struktur

Denne analysen er bygd opp av sju kapitler. I kapittel 1 har oppgaven blitt presentert med en introduksjon og forklaring av problemstillingen. Kapittel 2 tar for seg relevant teori for analysen, ved å presentere tidligere litteratur om emnet. Oppgavens bidrag til litteraturen blir deretter introdusert. Kapittel 3 omhandler datamaterialet. Her vil oppgavens datagrunnlag og definisjon av urbane og rurale områder bli presentert. Videre utføres deskriptive analyser, og det kommenteres for styrker og svakheter ved datagrunnlaget. Kapittel 4 tar for seg valgt metode, ved å presentere minste kvadraters metode, samt enkel og multippel lineær regresjonsmodell. Antakelser for regresjonsmodellene, determinasjonskoeffisienten og hypotesetesting er videre beskrevet i kapittelet. Forklaring av variabler og valg av funksjonsform introduseres i kapittel 5, før regresjonsanalysene bli utført i samme kapittel. Resultatene av disse analysene presenteres tabellarisk, og det blir utført en analyse for å teste robusthet. Videre blir resultatene tolket og diskutert i kapittel 6, med følgende drøfting av feilkilder. Det gis også en anbefaling til videre forskning som en utvidelse av oppgavens analyse. Avslutningsvis vil den empiriske analysen konkluderes i kapittel 7.

2. Teoretisk rammeverk

I dette kapitlet presenteres tidligere empiri og relevant teori for å sette den teoretiske rammen for oppgaven. For å danne en dypere forståelse av de underliggende årsakene til lønnsforskjellene i urbane og rurale områder i Norge, introduseres litteratur som tar for seg lønnsgap. Deretter vil oppgavens bidrag til denne litteraturen bli presentert. Tidligere litteratur blir dermed sett i sammenheng med oppgavens problemstilling.

2.1 Tidligere litteratur

Utarbeidelsen av problemstillingen for oppgaven er forankret i vinklinger og funn ved eksisterende empiri og teori om lønnsgap mellom urbane og rurale områder. Det eksisterer en omfattende mengde forskning på feltet, og litteraturen presentert i dette delkapitlet er utvalgt for å gi et innblikk i ulike tilnærminger og forklaringer til bakgrunnen for lønnsgapet.

Hovedgrunnlaget for den teoretiske rammen i oppgaven er presentert ved to artikler av de la Roca og Puga (2017) og Dauth et al. (2022), som konkluderer med ulike aspekter som påvirker lønnsgapet mellom urbane og rurale områder. Studiene er basert på internasjonale forhold, henholdsvis Spania og Tyskland. Forskning av Carlsen et al. (2016) og Leknes et al. (2022) bygger videre på disse artiklene, og tar utgangspunkt i data om Norges arbeidsstyrke, som er svært aktuelt med tanke på ønsket om å finne forklaringer for lønnsgapet i Norge. Ved å inkludere utdanning i analysene anses de to sistnevnte artiklene som ytterligere relevante, da det kan være interessant å se på avkastningen av en investering i utdanning.

Bylønnspremie

I store deler av forskningen om lønn i urbane og rurale områder er det bred enighet om at det eksisterer en bylønnspremie. Det fremkommer av forskningen til de la Roca & Puga (2017) at lønnen i spanske byer er høyere enn i mindre byer og rurale områder i Spania. Forskerne viser til at dette også er gjeldende for de fleste andre utviklede land. Selv ved sammenligning av arbeidstakere i samme bransje, med lik utdanning og arbeidserfaring, vil bylønnspremien være av betydelig verdi, ifølge artikkelen. På en annen side er også kostnadene høyere i byer. For

arbeidstakere innebærer dette blant annet økte leie- og boligpriser, som resulterer i at de må kompenseres med høyere lønn. Økte kostnader forbundet med drift i urbane områder gjør det nødvendig for bedrifter å oppnå produktivitetsfordeler, for at det skal være lønnsomt å holde til i byen. De la Roca og Puga (2017) kommer frem til tre faktorer som påvirker bedrifters produktivitet i byene, og som dermed gjør dem villig til å betale høyere lønn til sine ansatte. Den første faktoren forskerne vektlegger er statiske agglomerasjonseffekter, som innebærer fordeler som kun finnes i byene, og vil gå tapt dersom man flytter fra byen. Videre forklares en sorteringstendens, som i denne sammenheng går ut på at mer produktive arbeidstakere ofte velger å holde til i byen. Denne faktoren blir videre diskutert i avsnittet om assortativ matching og bystørrelse. Avslutningsvis vektlegger de la Roca og Puga (2017) at arbeidstakere oppnår verdifull læring og erfaring, på grunn av de mange mulighetene som finnes i byene. Denne fordelene gir økt humankapital og følger arbeidstakerne gjennom hele livet.

Det er også foretatt forskning om bylønnspremien på norske forhold. Carlsen et al. (2016) utvider analysen til de la Roca og Puga (2017), og setter søkelys på bylønnspremien for de største byene i Norge, definert med et innbyggertall over 150 000. Artikkelen tar utgangspunkt i data fra 1993-2010 som omhandler Norges arbeidsstyrke i privat sektor. Analysen fremhever at bylønnspremien er større for arbeidstakere med høyere utdanning, noe som øker med arbeidserfaring. Videre øker bylønnspremien for høyt utdannede som bytter arbeidsgiver, mens den for lavere utdannede heller øker med tiden de jobber i én bedrift (Carlsen et. al., 2016), som et resultat av opparbeidet ansiennitet.

Assortativ matching og bystørrelse

Bylønnspremien kan også forklares ved å analysere antallet tilgjengelige arbeidstakere- og givere, samt kvaliteten på disse partene. Ifølge en studie basert på privat sektor i Tyskland, har assortativ matching betydning for geografiske lønnsforskjeller (Dauth et al., 2022). Studiet forklarer begrepet assortativ matching ved at høyt kvalifiserte arbeidstakere har en tendens til å holde til i byer hvor mange bedrifter anses å være av høy kvalitet, en tendens også beskrevet av de la Roca og Puga (2017). Ifølge Dauth et al. (2022) omfatter begrepet assortativ matching ytterligere at det høye antallet arbeidstakere og arbeidsgivere i byene fører til at byene er mer

effektive i matching av disse. Oppsummerende forklarer assortativ matching at bedrifter og arbeidstakere av høy kvalitet er overrepresentert i byene, og vil i større grad og mer effektivt bli matchet basert på lokalisering. Som følge av dette øker produktiviteten og lønnen i byene. Til sammenligning vil tilbud og etterspørsel av arbeidstakere og arbeidsgivere i mindre grad samsvare med kvalitet og kvantitet av disse i mindre byer og rurale områder (Dauth et al., 2022), hvilket resulterer i et lønnsgap mellom urbane og rurale områder. Bylønnspremien kan dermed forklares på bakgrunn av de to faktorene som inngår i assortativ matching.

Forholdet mellom assortativ matching og bystørrelse er også analysert for Norges arbeidsstyrke, med utgangspunkt i data som omfatter privat sektor fra 2003-2014. Leknes et al. (2022) bygger videre på studiet av Dauth et al. (2022), og inkluderer utdanningsnivå i analysen. Det fremkommer at høyt utdannede arbeidstakere driver en positiv sammenheng mellom bystørrelse og assortativ matching. Sammenhengen er positiv for høyt utdannede uavhengig av kjønn og alder, men i ulik størrelsesorden. Ytterligere belyses en positiv påvirkning av bystørrelse for kvinner med videregående skole som høyeste fullførte utdanning (Leknes et al., 2022). Dette indikerer at kvinner i større grad drar nytte av økt bystørrelse, da de blir bedre matchet i byer uavhengig av om de fullfører høyere utdanning enn videregående skole.

2.2 Oppgavens bidrag til litteraturen

Det teoretiske fundamentet poengterer relevansen av alder, kjønn og utdanning for geografiske lønnsforskjeller. Disse variablene benyttes også i denne oppgaven, men beregnes på et annet grunnlag. Ved tidligere litteratur har hovedfokuset vært på privat sektor, og et skille mellom urbane og rurale områder som falt på bakgrunn av eksempelvis bystatus og innbyggertall.

Oppgavens bidrag til litteraturen kommer til uttrykk på bakgrunn av ønsket om en mer nyansert analyse, og det anses derfor som hensiktsmessig å se på det totale arbeidsmarkedet. Som følge av dette, inkluderes både offentlig og privat sektor, da det kan tenkes at det totale arbeidsmarkedet i rurale områder består av en større andel offentlige arbeidsplasser, sammenlignet med urbane områder. Videre utvides det nyanserte bildet av lønnsforskjeller ved bruk av sentralitetsindeksen som mål for urbane og rurale områder.

Artiklene i litteraturgjennomgangen tar i bruk paneldata i sine empiriske analyser. For å gi svar på oppgavens problemstilling vurderes derimot tverrsnittsdata som mest hensiktsmessig, da det er lønnsforskjeller i dagens samfunn som ønskes analysert. Nyere tverrsnittsdata om sentralitet og demografiske forskjeller anvendes som utgangspunkt for analysen, hvilket bidrar til et oppdatert bilde av lønnsforskjeller i Norge i 2022. Det er rimelig å anta at endringer i et lønnsgap mellom urbane og rurale områder ikke er betydelig ved en kortere tidsperiode, da lønnsnivået er forholdsvis stabilt. Ved bruk av paneldata ville det følgelig vært nødvendig å inkludere observasjoner over flere tiår for å gi utslagsgivende resultater av betydning for analysen, utover resultatene som fremkommer av tverrsnittsdataen. Data fra en større tidsperiode ville økt risikoen for feilmargin gjennom at uttrekket inneholder radikale endringer og tendenser i arbeidsmarkedet, som problemstillingen ikke har til hensikt å belyse. Et eksempel her kan være den positive økningen av kvinnelig deltakelse i yrkeslivet, hvilket ikke omfattes av oppgavens problemstilling.

3. Datamateriale

I dette kapitlet presenteres datamaterialet som benyttes i regresjonsmodellene i kapittel 5. Innledningsvis beskrives valg av datagrunnlag, før definisjonen av urbane og rurale områder blir fremlagt. Deskriptive analyser blir presentert, herunder tabellarisk fremstilling av deskriptiv statistikk for både den avhengige variabelen og de uavhengige variablene. Avslutningsvis blir styrker og svakheter ved datasettet poengtert.

3.1 Datagrunnlag

I denne oppgaven benyttes statistikkbanken til SSB for å danne datagrunnlaget for analysen. I statistikkbanken er det tatt utgangspunkt i tre tabeller. Det er kun hentet ut data fra år 2022 i tabellene, siden det er det nyeste året i tabell 09429.

Den første tabellen er tabell 12852: Kommunefordelt månedslønn, etter bosted, arbeidssted, alder og kjønn (K) 2015 –2023 (Statistisk sentralbyrå [SSB], 2024). Tabellen filtreres for å kun gi informasjon som er relevant for oppgaven. Månedslønn er valgt som den forklarte statistikkvariabelen og gjennomsnitt som statistikk mål. Tabellen skiller månedslønnen i Norges kommuner ut ifra arbeidstakernes alder og kjønn.

Den andre tabellen datagrunnlaget baseres på, inneholder data om utdanningsnivå, og er tabell 09429: Utdanningsnivå, etter kommune og kjønn (K) 1970 – 2022 (SSB, 2023). Denne filtreres for å få opplyst antall mennesker, herunder også fordelt på kjønn, med ulik utdanning fordelt på Norges kommuner. Dataen kategoriseres etter høyeste fullførte utdanningsnivå, ved grunnskole, videregående skole, fagskole, kort høyere utdanning og lang høyere utdanning. Skillet mellom kort og lang høyere utdanning er 4 år, for å separere bachelorgrad fra høyere utdanningsnivå (SSB, 2023).

Den siste tabellen som benyttes er tabell 07459: Alders- og kjønnsfordeling i kommuner, fylker og hele landets befolkning (K) 1986-2024 (SSB, 2024). Tabellen benyttes for å hente ut antall

mennesker for alle Norges kommuner fordelt på følgende aldersgrupper: under 24 år, 25-34 år, 35-44 år, 45-69 år, 70-79 år og 80 år eller eldre.

3.2 Definisjon av urbane og rurale områder

For å være i stand til å svare på oppgavens problemstilling, må det defineres hvilke kommuner som klassifiseres som urbane og rurale i Norge. Det finnes mange definisjoner og utgangspunkt for å danne et slikt skille. Eksempelvis benyttet Carlsen et al. (2016) et innbyggertall på over 150 000, som presentert i kapittel 2.1. Andre mål på urbanitet kan være bystatus eller befolkningstetthet. I denne oppgaven benyttes SSBs sentralitetsindeks som verktøy for å skille landets kommuner som urbane og rurale områder. Denne beslutningen er tatt på bakgrunn av at indeksen anses som mer nyansert og overveid sammenlignet med de ovennevnte definisjonene, som kan oppfattes å være for ensidige til å svare på problemstillingen. Sentralitetsindeksen anses som ytterligere hensiktsmessig å benytte som verktøy ved utviklingen av en definisjon, med tanke på at den er kompatibel med datagrunnlagets kommunefordeling.

Sentralitetsindeksen er basert på to delindekser; *“Hvor mange arbeidsplasser kan de som bor i hver enkelt grunnkrets nå med bil i løpet av 90 minutter”* (Høydahl, 2020, s. 7) og *“Hvor mange servicefunksjoner kan de som bor i hver enkelt grunnkrets nå med bil i løpet av 90 minutter”* (Høydahl, 2020, s. 7). Det er hensyntatt alle landets 13 500 bebodde grunnkretser, og disse er vektet etter antall innbyggere (Høydahl, 2020). I tillegg blir antall arbeidsplasser og servicefunksjoner vektet forskjellig, ved at større avstand har en kostnad og utfallsgivende negativ effekt på graden av sentralitet (Høydahl, 2020).

Indeksen deler landets kommuner inn i nivåer fra 1 til 6, der nivå 1 betegner de mest sentrale kommunene og nivå 6 de minst sentrale kommunene (Høydahl, 2020). I denne oppgaven er det besluttet å benytte nivå 1 og nivå 2 for å definere urbane områder, og nivå 3 til nivå 6 til å definere rurale områder. Dette gjøres fordi det anses som mest hensiktsmessig å ta utgangspunkt i ekstreme verdier for å svare på problemstillingen. Oslo, Bærum og Trondheim er eksempler på kommuner som befinner seg i nivå 1 eller 2, og er urbane områder etter oppgavens definisjon. Eksempler på kommuner i nivå 3 er Tromsø, Ålesund og Kristiansand. Til tross for at sistnevnte

kommuner har bystatus og relativt høye innbyggertall sammenlignet med kommuner i nivå 4, som eksempelvis Molde, Steinkjer og Narvik, kan det tenkes at disse byene ikke anses som spesielt attraktive for potensielle arbeidstakere. Dette argumentet gir grunnlag for å definere nivå 1 og 2 som urbane områder, og 3 til 6 som rurale områder.

I 2020 publiserte SSB tall og tilhørende notater for sentralitetsindeksen. I ettertid har de oppdatert sentralitetsindeksen med kommuner fra år 2024, basert på endringer i befolkningsdata per 1. januar 2023 (Høydahl, 2023). Siden denne oppgaven bygger på data fra år 2022, noe som begrunnes i kapittel 3.1, vil sentralitetsindeksen for 2020 benyttes. På denne måten vil kommunene for sentralitetsindeksen og fra datamaterialet være de samme. Dette gir oppgaven et nøyaktig grunnlag for å sammenligne landets urbane og rurale områder.

3.3 Deskriptive analyser

Deskriptiv statistikk er en form for oppsummering og presentasjon av statistisk informasjon som skal være enkel for leseren å tolke (Anderson et al, 2020, s. 13). I denne oppgaven utformes deskriptiv statistikk for både den avhengige variabelen og de uavhengige variablene. Bakgrunnen for dette er at det i kapittel 5 skal utføres sammensatte og komplekse regresjonsanalyser med variablene, og at det derfor kan anses som nyttig med en tabellarisk utforming av dataene som skal benyttes videre i regresjonsanalysen.

Deskriptiv statistikk for avhengig variabel

Opgavens avhengige variabel er gjennomsnittlig månedslønn. Ved å klassifisere Norges kommuner som urbane og rurale ved anvendelse av sentralitetsindeksen, danner dette grunnlag for å beregne gjennomsnittlig månedslønn for kommunene fordelt på de to kategoriene. Tabellen under illustrerer gjennomsnittlig månedslønn i urbane og rurale områder, og oppgir også tilhørende standardavvik. I tillegg oppgir tabellen antall observasjoner, samt minimum og maksimum verdi ved disse observasjonene.

	Urbane områder	Rurale områder
Gjennomsnittlig månedslønn	52 214,40	47 154,83
Minimum	46 460,00	39 350,00
Maksimum	64 810,00	64 470,00
Standardavvik	4 270,49	2 566,57
Antall observasjoner	25	331

Tabell 1: Gjennomsnittlig månedslønn i urbane og rurale områder. Med unntak av antall observasjoner er alle tall oppgitt i kroner og avrundet til to desimaler.

Som illustrert i Tabell 1, klassifiseres 25 kommuner som urbane, og 331 kommuner som rurale, noe som til sammen utgjør Norges 356 kommuner per 2022. I urbane områder var gjennomsnittslønnen 52 214,40 kroner, mens den var 47 154,83 kroner i rurale områder. Dette utgjør en differanse på 5 059,57 kroner. Med andre ord er det et tydelig lønnsgap mellom urbane og rurale områder, noe som underbygger resultatene av tidligere empiri om bylønnspremie.

Blant de urbane kommunene var minste gjennomsnittlig månedslønn 46 460,00 kroner, mens maksimum var på 64 810,00 kroner. De rurale kommunene viser seg å ha minimum gjennomsnittlig månedslønn på 39 350,00 kroner, og maksimum var på 64 470,00 kroner. Blant de urbane kommunene, er det Bærum som har høyest gjennomsnittlig månedslønn, mens det er Gjerdrum som har den laveste. Sola er den kommunen som har høyest gjennomsnittlig månedslønn blant de rurale kommunene, og Frosta har den laveste.

Standardavviket for urbane kommuner er 4 270,49, mens standardavviket for rurale kommuner er 2 566,57. Standardavviket for urbane kommuner er dermed over dobbelt så høyt som for de rurale kommunene. Standardavviket for de urbane kommunene indikerer derfor en større variasjon rundt gjennomsnittslønnen, noe som kan forklares av loven om store tall. Denne teorien indikerer at det er nødvendig at et utvalg består av et tilstrekkelig høyt antall observasjoner for å være representativt for populasjonen (Wooldridge, 2020, s. 722). Det lave antallet av observasjoner i urbane områder relativt til observasjonene i rurale, betyr at observasjonene for urbane områder muligens ikke er like representative for tilhørende gjennomsnitt.

Deskriptiv statistikk for uavhengige variabler

Opgavens uavhengige variabler består av geografisk område, alder, kjønn og høyeste grad av fullførte utdanning. Den deskriptive statistikken for disse variablene inndeles i to tabeller for en oversiktlig fremstilling.

Tabell 2 illustrerer deskriptiv statistikk for høyeste grad av fullførte utdanning. Oppgavens datagrunnlag for utdanningsnivå uttrykker antall mennesker som har ulik grad av høyeste fullførte utdanning i Norges kommuner. Den oversiktlige presentasjonen som fremkommer ved deskriptiv analyse av utdanning vil være nyttig informasjon ved tolkning av regresjonsanalysene som gjennomføres i kapittel 5.

	Urbane områder	Rurale områder
Grunnskole	23,09%	27,45%
Videregående skole	34,56%	42,82%
Fagskole	2,94%	3,51%
Mindre enn 4 år på høyskole eller universitet	26,74%	20,50%
Mer enn 4 år på høyskole eller universitet	12,67%	5,72%

Tabell 2: Gjennomsnittlig andel av befolkningen i urbane og rurale områder med høyeste fullførte grad av utdanning.

Det kommer frem av Tabell 2 at en gjennomsnittlig større andel av befolkningen i rurale områder, relativt til urbane, har grunnskole, videregående skole og fagskole som høyeste fullførte grad av utdanning. På den andre siden har en høyere andel av befolkningen i urbane områder tatt høyere utdanning. Det er verdt å nevne at det gjennomsnittlig i urbane kommuner er 12,67% av befolkningen som har en høyere utdanning på over 4 år. I rurale områder er denne andelen 5,72%. Denne observasjonen indikerer at en større andel høyt utdannede befinner seg i urbane områder.

Tabell 3 illustrer gjennomsnittlig månedslønn med hensyn på de uavhengige variablene alder og kjønn oppgitt i kroner. Også for disse uavhengige variablene vil det være nyttig å utføre en

deskriptiv analyse for å illustrere de eksisterende lønnsforskjellene før regresjonsanalysen utføres.

	Urbane områder	Rurale områder	Differanse
Under 30 år	39 023,60	38 270,68	752,92
Mellom 30-54 år	54 147,60	48 713,65	5 433,95
55 år og eldre	58 322,80	50 573,42	7 749,38
Mann	55 129,20	48 855,60	6 273,60
Kvinne	48 497,20	44 927,05	3 570,15

Tabell 3: Gjennomsnittlig månedslønn med hensyn på alder og kjønn i urbane og rurale områder. Alle tall oppgitt i kroner og avrundet til to desimaler.

Det er gjennomgående i Tabell 3 at gjennomsnittlig månedslønn er høyere i urbane områder enn i rurale, uavhengig av aldersgruppe. Differansen mellom gjennomsnittlig lønn i urbane og rurale områder for de under 30 år er på kun 752,92 kroner. Differansen øker til 5 433,95 kroner for de mellom 30-54 år, og øker ytterligere til 7 749,38 kroner for de over 50 år. Den økende differansen underbygger at økende alder kan være en driver til lønnsgapet mellom urbane og rurale områder. I likhet med alder er gjennomsnittlig månedslønn høyere i urbane områder enn i rurale områder, uansett kjønn. Bylønnspremien er 6 273,60 kroner for menn og 3 570,15 kroner for kvinner. Observasjonen illustrerer med andre ord at menn i større grad kan bidra til lønnsforskjeller mellom urbane og rurale områder.

3.4 Styrker og svakheter ved datagrunnlaget

Datasettene fra SSB inneholder flere momenter som kan betraktes som styrker i sammenheng med denne oppgaven. Først og fremst inneholder datasettene flere variabler som kan bidra til å forklare forskjellen mellom gjennomsnittslønnen i urbane og rurale områder i Norge. Variablene for geografisk område, alder, utdanning og kjønn vil likevel ikke være i stand til å ta for seg alle forhold som forklarer gjennomsnittslønnen, noe som vil bli fanget opp av et feilledd. Et mer omfattende datasett med flere uavhengige variabler ville ført til at en enda større del av bylønnspremien kunne blitt forklart av modellene, slik at feilleddet minsker ytterligere. I lys av

tidligere litteratur og valgt problemstilling er det antatt at de utvalgte demografiske faktorene for denne oppgaven i stor grad vil være forklarende for bylønnspremien i Norge. På denne måten blir ikke fåtallet av uavhengige variabler sett på som en svakhet, men heller som en styrke, da det i datasettene foreligger detaljert informasjon innen de fire uavhengige hovedvariablene for geografisk område, utdanning, alder og kjønn. Det kan også anses som en styrke at datasettet skiller mellom ulike utdanningsnivå, siden det kan tenkes at grad av utdanning er av større betydning for avkastning på utdanning, enn hvor mange år personen har gått på skole. Avslutningsvis styrker den store populasjonen i datasettet sannsynligheten for at resultatene er generaliserbare for hele Norges arbeidsstyrke.

På en annen side inneholder datasettet noen svakheter, med at det ikke hensyntar erfaring eller skillet mellom offentlig og privat sektor. Det hadde vært interessant å se effekten av disse på bylønnspremien og feilledet, da det teoretiske fundamentet fant at disse variablene var av betydning. Videre er det rimelig å anta at forskjellige bransjer og yrker, samt ulike utdanninger med samme tidsløp, vil ha varierende avkastning på gjennomsnittlig lønn. Heller ikke dette gir datasettet informasjon om, og dermed ikke muligheten til å utforske. Videre er det relevant å nevne at datasettet mangler tall for enkelte kommuner i rurale områder. Dette kan føre til at analysen ikke er fullstendig representativ for populasjonen. Denne svakheten betraktes likevel ikke som betydelig, noe som diskuteres nærmere under MLR-antakelsene i delkapittel 4.3. Avslutningsvis gir datasettet ingen mulighet til å se på marginaleffekten av ett års ekstra utdanning eller ett år høyere alder. Av disse kan det tenkes at muligheten til å se på marginaleffekten for alder ville vært av størst nytte. Skillet mellom aldergrupper er så stort i disse datasettene at det vil være vanskelig å beregne et vendepunkt som forklarer hvilken alder som har størst påvirkning for gjennomsnittslønnen, og når denne trenden snur. En ytterligere svakhet angående alder innebærer at Tabell 09429 ikke inneholder lønnsdata, men kun benytter antall mennesker som statistikkvariabel. Til sammenligning inneholder det øvrige datagrunnlaget tall for lønn, hvor aldersfordelingen i tillegg er distribuert ved andre og færre aldersintervaller.

4. Metode

Dette kapitlet presenterer metoden som benyttes for å besvare problemstillingen. Først presenteres minste kvadraters metode og hvorfor denne er nyttig for oppgaven. Videre forklares både enkel og multipl linear regresjonsmodell og hvordan modellene benyttes. Det finnes seks ulike antakelser som må holde for at modellene skal være uten bias, og være gyldige for å kunne benyttes i hypotesetesting, en metode som benyttes i kapittel 5 for å teste svarene gitt i regresjonsanalysene. I tillegg vil determinasjonskoeffisienten bli forklart.

4.1 Minste kvadraters metode (MKM)

Minste kvadraters metode (MKM) benyttes for å estimere koeffisientene i regresjonsmodellen, ved å finne koeffisientene som minimerer summen av de kvadrerte anslåtte avvikene (Sucarrat, 2017, s. 56). Med andre ord minimerer MKM variansen for modellen. Beregningene blir ikke gjort manuelt, men metoden benyttes ved å utføre regresjonsanalyser i programmet Stata.

4.2 Enkel og multipl linear regresjonsmodell (SLR og MLR)

For å analysere lønnsforskjeller mellom urbane og rurale områder i Norge benyttes linear regresjon av to typer, enkel og multipl. Disse gir mulighet for å analysere sammenhengen mellom en avhengig og en eller flere uavhengige variabler. Denne oppgaven tar for seg lønn som avhengig variabel, og ønsker å studere hvordan de uavhengige variablene urbant eller ruralt område, utdanning, alder og kjønn påvirker lønn i urbane og rurale områder. Den lineære effekten av de uavhengige variablene på den avhengige variabelen kan estimeres gjennom de lineære regresjonsmodellene (Wooldridge, 2020, s. 804).

Enkel linear regresjonsmodell er et ofte brukt begrep i økonometrien og forkortes til SLR (fra engelsk: simple linear regression). SLR-modellen inneholder én avhengig variabel, som er en linear funksjon av kun én uavhengig variabel og ett feilledd (Wooldridge, 2020, s. 809). Det kan være nyttig å benytte SLR for å undersøke den isolerte effekten den uavhengige variabelen har på den avhengige variabelen. Multipl linear regresjonsmodell (MLR) har i likhet med SLR-

modellen én avhengig variabel, men denne uttrykkes som en funksjon av flere uavhengige variabler i tillegg til feilledet (Wooldridge, 2020, s. 805). Med andre ord gir en MLR-modell mulighet til å se på sammenhengen mellom lønn og flere uavhengige variabler samtidig, og gir dermed et mer helhetlig bilde på hvordan de uavhengige variablene påvirker lønnsgapet (Sucarrat, 2017, s. 14).

Denne empiriske analysens modeller blir oppgitt først i delkapittel 5.2. Deretter, blir funksjonsformen for både SLR- og MLR-modeller beskrevet i delkapittel 5.3. Dette for å få sett valgte funksjonsformer i sammenheng med de gjeldende modellene.

4.3 MLR-antakelser

MKM-modellens gyldighet og pålitelighet sikres ved at MLR-antakelsene innfris. Det er grunnlag for å påstå at MKM estimatorene er uten bias ved at MLR. 1-4 holder (Wooldridge, 2020, s. 79). MLR. 5 og MLR. 6 tar for seg forhold som varians og standardavvik, noe som benyttes i hypotesetesting. Ved å holde antakelsene MLR. 5 og MLR. 6, kan man derfor anta å beregne gyldige hypotesetester. I MLR-antakelsene diskuteres det hvorvidt kausaleffekten av en uavhengig variabel på den avhengige variabelen er funnet (Wooldridge, 2020, s. 798). Det vil si effekten på gjennomsnittslønn som følge av en endring av urbant eller ruralt område, utdanning, alder eller kjønn, gitt at alt annet er likt. I dette delkapittelet presenteres derfor de ulike antakelsene før MLR-modellene beregnes i kapittel 5.

MLR. 1

MLR. 1 definerer den multiple lineære modellen, og handler om at koeffisientene i modellen må ha perfekt linearitet (Wooldridge, 2020, s. 79-80). Modellene i denne oppgaven inneholder kun lineære koeffisienter, og det antas dermed at denne antakelsen holder i alle tilfeller.

MLR. 2

MLR. 2 er antakelsen om et tilfeldig utvalg i modellen (Wooldridge, 2020, s. 80). I noen av kommunene er det ikke registrerte tall for alle de ulike variablene. Det er gjentakende at det er

små kommuner med lav grad av sentralitet som mangler ulik data. Denne mangelen kan tenkes å skyldes ulike faktorer, eksempelvis mangel på nok innbyggere til å finne data. På bakgrunn av det høye antallet observasjoner i kategorien rurale kommuner, vil datamangel i et fåtall av disse kommunene antakeligvis ikke påvirke resultatet i betydelig grad. Dette kan underbygges av teorien om store talls lov. På bakgrunn av disse argumentene antas det at MLR. 2 holder.

MLR. 3

MLR. 3 tar for seg antakelsen om ingen perfekt kollinearitet. Poenget i denne antakelsen er at de uavhengige variablene kan korrelere, men at en eventuell korrelasjon ikke kan være perfekt. Dersom man ikke skulle tillatt korrelasjon mellom uavhengige variabler, ville nytten av MLR vært svært begrenset (Wooldridge, 2020, s. 80-81). I delkapittel 3.3 konstateres det at bylønnspremien øker med både alder og utdanning, men også med det å være mann. På bakgrunn av dette, kan det være naturlig å anta at disse uavhengige variablene også påvirker hverandre. På den andre siden vil det være naturlig å anta at disse forholdene ikke er perfekt korrelert. Et eksempel fra oppgaven er at alder og utdanning kan antas å ha korrelasjon. Jo eldre man blir, jo høyere sannsynlighet er det for at man har oppnådd en høyere grad av fullført utdanning. Disse variablene er derimot ikke perfekt korrelerte. Eksempelvis velger enkelte å ikke ta høyere utdanning og noen fullfører ikke utdanningen de har begynt på, samtidig med økende alder. Dette poengterer at de uavhengige variablene ikke er perfekt kollinear, og det antas derfor at MLR. 3 holder.

MLR. 4

MLR. 4 kan anses som den viktigste antakelsen, og omhandler forventningen om at verdien av feilledet i regresjonsmodellen skal være lik null, gitt de uavhengige variablene (Wooldridge, 2020, s. 811). De gitte uavhengige variablene i denne oppgaven; geografisk område, alder, kjønn og utdanning, vil sannsynligvis ikke danne et godt nok grunnlag for å kunne gi et helhetlig perspektiv på hvilke faktorer som kan påvirke lønnsgapet mellom urbane og rurale områder. Det vil høyst sannsynlig være andre faktorer som påvirker gapet i tillegg til disse tre uavhengige faktorene. Som ved alle andre forhold i en kompleks verden, vil det ofte være nærmest umulig å kartlegge alle uavhengige variabler i en slik analyse. I denne oppgaven er fokuset rettet mot å

isolert sett finne hvilken påvirkning de uavhengige variablene har på lønnsgapet i urbane og rurale områder. Feilleddet i modellen vil derfor antas å være lik null for å kunne isolere effektene.

MLR. 5

MLR. 5 omhandler antakelsen om homoskedastisitet. Denne antakelsen uttaler at variansen i feilleddet skal være konstant, gitt de uavhengige variablene (Wooldridge, 2020, s. 802). Det vil si at avviket fra regresjonslinjen for alle observasjonene skal være like. Det motsatte av homoskedastisitet er heteroskedastisitet. Heteroskedastisitet oppstår når standardavviket i en modell påvirkes av hvorvidt en dummyvariabel er lik null eller én (Sucarrat, 2017, s. 71). Ved heteroskedastisitet vil ikke hypotesetestene være gyldige. Grunnen til dette er at den beregnede verdien av variansen er feil, noe som kan føre til at man konkluderer uriktig i en hypotesetest. (Sucarrat, 2017, s. 136). Det antas derfor i denne oppgaven at MLR. 5 holder.

MLR. 6

Den siste antakelsen antar en normalfordeling av feilleddet (Wooldridge, 2020, s. 118). MLR. 6 presiserer også at normalfordelingen er fordelt rundt et gjennomsnitt på null og har en varians lik σ^2 (Sucarrat, 2017, s. 70). Ved å anta at MLR. 6 holder, vil t-testene være t-distribuert, samt at F-testene er F-distribuert (Sucarrat, 2017, s. 71). Dette danner grunnlag for å utføre hypotesetester med gyldige konklusjoner. Det antas derfor at MLR. 6 holder slik at hypotesetestene kan utføres.

4.4 Determinasjonskoeffisienten

Determinasjonskoeffisienten, R^2 , er et mål på hvor god forklaringskraft modellen har (Sucarrat, 2017, s. 58). Verdien på R^2 kan variere mellom 0 og 1, hvor 1 tilsier at modellen har en fullstendig forklaringskraft, mens en verdi på 0 tilsier ingen forklaringskraft (Sucarrat, 2017, s. 58). Etter hvert som flere uavhengige variabler legges til i en modell, vil R^2 øke på grunn av reduseringen dette har på feilleddet. For å unngå at forklaringskraften til R^2 skal bli overestimert,

er det vanlig å bruke en justert determinasjonskoeffisient (Anderson et al, 2020, s. 698). Det ses derfor bort fra R^2 oppgitt i vedlegg 2, og benytter heller den justerte R^2 utregnet i Stata. Denne justerte R^2 inkluderer antall observasjoner og antall uavhengige variabler for å få en mer nøyaktig verdi (Anderson et al, 2020, s. 699). I denne oppgaven er det derfor benyttet justert R^2 , slik at estimatet av determineringskoeffisienten blir mest mulig nøyaktig.

4.5 Hypotesetesting

For å fastslå om en uavhengig variabel har en påvirkning på den avhengige variabelen, kan dette undersøkes ved å teste ulike hypoteser. Den alternative hypotesen, H_A , vil være den hypotesen man ønsker å teste, mens nullhypotesen, H_0 , vil være alle andre utfall enn H_A (Sucarrat, 2017, s. 23). H_0 vil normalt sett være hypotesen om at det ikke finnes en sammenheng mellom variablene ($H_0: \beta_i = 0$), mens H_A vil være hypotesen om at det finnes en sammenheng ($H_A: \beta_i \neq 0$). Ved hjelp av hypotesetester kan man konkludere med å forkaste eller ikke forkaste H_0 (Sucarrat, 2017, s. 24). Det kan benyttes ulike typer hypotesetester, og det vil i denne oppgaven bli benyttet to ulike typer; t-test og F-test.

t-test

Ved å benytte en t-test vil man kunne teste sammenhengen mellom to variabler (Wooldridge, 2020, s. 801 og 810). En t-test beregnes ved følgende formel (Sucarrat, 2017, s. 61):

$$\frac{\beta - H_0 \text{ verdi}}{se(\beta)}$$

β er koeffisienten man ønsker å teste om har en effekt (Sucarrat, 2017, s. 60). H_0 verdi illustrerer verdien man ønsker å teste om β -en har gitt nullhypotesen. $se(\beta)$ står for standardavviket til β (Sucarrat, 2017, s. 61). Løsningen av denne testobservasjonen sammenlignes med en kritisk verdi i en tabell for t-distribusjon. For å kunne lese av denne tabellen, er det nødvendig å vite antall observasjoner, n , og antall uavhengige variabler, k . Disse må være kjent for å kunne finne antall frihetsgrader for testobservasjonen, noe som er nødvendig å vite for å kunne lese av

tabellen for t-verdier. Finner frihetsgrader ved $n - k$ (Sucarrat, 2017, s. 61). Ved å lese av for frihetsgrader i tillegg til signifikansnivå, finner man tilhørende kritisk verdi. Når man sammenligner den kritiske verdien med verdien for testobservasjonen, kan man konkludere med å forkaste eller ikke forkaste H_0 (Wooldridge, 2020, s. 799). Hovedregelen for en t-test er å forkaste H_0 dersom testobservasjonen er større enn den positive absoluttverdien av den kritiske verdien, eller mindre enn den negative absoluttverdien av den kritiske verdien (Sucarrat, 2017, s. 61).

F-test

Ved å benytte en F-test kan flere uavhengige variabler hypotesetestes samtidig ved å beregne to modeller (Sucarrat, 2017, s. 64-65). Den første modellen beregnes med restriksjonene gitt i H_0 , slik at de uavhengige variablene som ønskes testet er satt lik null. Den andre modellen beregnes uten restriksjonene. Når disse modellene er beregnet kan vi benytte formelen for F-testen (Sucarrat, 2017, s. 65):

$$F = \frac{(R_{ur}^2 - R_r^2)/m}{(1 - R_{ur}^2)/(n - k)}$$

Her er R_{ur}^2 lik som R^2 for modellen uten restriksjoner, mens R_r^2 er R^2 for modellen med restriksjoner. Antall påstander i H_0 er benevnt som m , og antall observasjoner er benevnt som n . Antallet av parametere som er beregnet for modellen uten restriksjoner er benevnt som k (Sucarrat, 2017, s. 65-66).

På samme måte som for t-testen skal det beregnede testuttrykket for F-testen sammenlignes med kritisk verdi, men her etter en kritisk verdi i tabell for F-distribusjon. For å finne tilhørende kritisk verdi, må det beregnes frihetsgrader for både teller og nevner i testuttrykket. Frihetsgrader for teller vil være lik m , mens frihetsgrader for nevner er lik $n-k$ (Sucarrat, 2017, s. 67). Om testverdien befinner seg innen forkastningsområdet, altså er større eller lik kritisk verdi, forkastes H_0 (Sucarrat, 2017, s. 67).

5. Regresjonsanalyser

I dette kapittelet utføres oppgavens regresjonsanalyser. Regresjonsanalysene er delt opp i tre hovedmodeller, hvor Modell 1 og Modell 2 har to submodeller hver. Den samlede analysen inneholder mange uavhengige variabler, noe som gjør det hensiktsmessig og ryddigere å dele opp i flere modeller. Under hver modell skal løsningen av modellen diskuteres, og det vil utføres tester for å undersøke ulike koeffisienters statistiske signifikans. Avslutningsvis vil det utføres en robusthetsanalyse av regresjonsanalysene.

5.1 Forklaring av variabler

På bakgrunn av at regresjonsanalysene består av flere modeller med ulike variabler anses det som hensiktsmessig med en grundig forklaring av disse. Modell 1.1 har gjennomsnittlig månedslønn som avhengig variabel, mens resten av modellene har logaritmen av gjennomsnittlig månedslønn som avhengig variabel. Alle modellene inneholder dummyvariabelen *urbant*, hvor variabelen har verdien 1 om kommunen klassifiseres som urban, og 0 om den klassifiseres som rural. Resten av de uavhengige variablene vil være unike for hver enkelt modell. Fullstendige notasjoner og forklaringer for hver enkelt variabel og koeffisient er formulert i vedlegg 1.

5.2 Modellene

Modell (1) ønsker primært å undersøke om det er grunnlag for å påstå at det er lønnsforskjeller i urbane og rurale områder i Norge. Modell (2) bygger på modell (1), men utvider modellen ved å inkludere uavhengige variabler for grad av fullført utdanning, også fordelt på kjønn. Modell (3) benytter også modell (1) som utgangspunkt, men inkluderer i tillegg uavhengige variabler for intervaller av en kommunes befolkningsdistribusjon. I utformingen av modell (2) og (3), er det valgt å utelate variablene andel grunnskole og andel under 24 år. Ved å utelate en variabel i en gruppe, her for utdanning og alder, er hensikten å ikke havne i dummyvariabelfellen (på engelsk: *dummy variable trap*). Denne fellen går ut på å inkludere for mange dummyvariabler i modellen (Wooldridge, 2020, s. 223).

Modell (1): Bylønnspremie

$$\text{Modell (1.1): } gj_lønn = \beta_0 + \beta_1 * \text{urbant} + u$$

$$\text{Modell (1.2): } \log_lønn = \beta_0 + \beta_1 * \text{urbant} + u$$

Modell (2): Utdanning og kjønn

$$\text{Modell (2.1): } \log_lønn = \beta_0 + \beta_1 * \text{urbant} + \beta_2 * \text{andel_vgs} + \beta_3 * \text{andel_fagsk} + \beta_4 * \text{andel_høyereutd_kort} + \beta_5 * \text{andel_høyereutd_lang} + u$$

$$\text{Modell (2.2): } \log_lønn = \beta_0 + \beta_1 * \text{urbant} + \beta_2 * \text{andel_vgs_kvinner} + \beta_3 * \text{andel_fagsk_kvinner} + \beta_4 * \text{andel_høyereutd_kort_kvinner} + \beta_5 * \text{andel_høyereutd_lang_kvinner} + \beta_6 * \text{andel_vgs_menn} + \beta_7 * \text{andel_fagsk_menn} + \beta_8 * \text{andel_høyereutd_kort_menn} + \beta_9 * \text{andel_høyereutd_lang_menn} + u$$

Modell (3): Alder

$$\text{Modell (3): } \log_lønn = \beta_0 + \beta_1 * \text{urbant} + \beta_2 * \text{andel_25til34år} + \beta_3 * \text{andel_35til44år} + \beta_4 * \text{andel_45til69år} + \beta_5 * \text{andel_70til79år} + \beta_6 * \text{andel_80ogeldre} + u$$

5.3 Valg av funksjonsform

Modell 1 er en SLR-modell, mens modell 2 og 3 er MLR-modeller. Forskjellen vil være at SLR-modellene har kun én uavhengig variabel, i motsetning MLR-modellene som har flere (Sucarrat, 2017, s. 54). Y_i er den avhengige variabelen og er følgelig den variabelen som ønskes estimert gitt β og X . X -ene er modellenes uavhengige variabler. β -ene kalles koeffisienter. β_0 illustrerer modellens konstantledd, altså hva Y vil bli gitt at alle X -verdiene er lik null. De andre β -ene er stigningstall og tolkes som oftest som hva én enhets endring i X påvirker den gjennomsnittlige endringen i Y , gitt at de andre X -ene er lik null. Til slutt inneholder begge formlene et feilledd,

ui. Feilleddet kan tolkes slik at den fanger opp feilen i predikasjonen til modellen (Sucarrat, 2017, s. 53-55).

Det er av betydning å nevne at modell (1.1) er en level-level modell. Ved å benytte en level-level modell i den første modellen beregnes bylønnspremien oppgitt i kroner. Det anses som hensiktsmessig å videre benytte funksjonsformen log-level. I en log-level modell er Y lik logaritmen av den avhengige variabelen, mens de uavhengige variablene vil være like som i en level-level modell (Sucarrat, 2017, s. 88). Ved å endre funksjonsform, vil de uavhengige variablene få en ny tolkning, da koeffisientene må multipliseres med 100% for å illustrere riktig effekt (Sucarrat, 2017, s. 88). For å gå fra en level-level modell til en log-level modell, er det tatt logaritmen av den avhengige variabelen, gjennomsnittlig månedslønn. Hensikten med å benytte logaritmen av gjennomsnittlig månedslønn er å se på en prosentvis endring i lønn. Dette gjøres for å kunne illustrere hvor mye de ulike uavhengige variablene øker eller reduserer gjennomsnittlig månedslønn. Dette vil også være fordelaktig om oppgaven senere ønskes benyttet til å sammenligne funn i andre land. I et slikt tilfelle vil ikke variabler oppgitt i kroner være et godt grunnlag for å sammenligne med land med annen valuta. Inflasjon er et annet argument for å benytte log-level modellen, ved at effekten av de uavhengige variablene på den avhengige variabelen kan sammenlignes med nye funn på et senere tidspunkt, til tross for at pengenes verdi vil endres over tid.

5.4 Modell (1): Bylønnspremie

Model (1.1): Bylønnspremie i kroner

Den avhengige variabelen i modell (1.1) er en kommunes gjennomsnittlige månedslønn, mens den uavhengige variabelen er en dummyvariabel for om kommunen klassifiseres som urban eller ikke. Verdien på koeffisienten som tilhører dummyvariabelen (β_1) kan ha to mulige utfall; 1 dersom kommunen kategoriseres som urban, og 0 dersom den kategoriseres som rural. Med andre ord er dette en SLR-modell hvor forholdet mellom gjennomsnittlig månedslønn og arbeidssted, urbant eller ruralt, blir estimert. Løsning av modellen blir:

$$\text{gjlønn} = 47\,154,83 + 5\,059,57 * \text{urbant}$$

Regresjonen viser at konstantleddet β_0 er 47 154,83. Denne verdien forklarer at man kan forvente at gjennomsnittslønnen i rurale kommuner er 47 154,83 kroner, gitt at dummyen for urbant er lik 0. β_1 er 5 059,57, hvilket betyr at det kan forventes at den gjennomsnittlige månedslønnen øker med 5 059,57 kroner om man betrakter en urban kommune. Forventet gjennomsnittlig månedslønn i urbane områder blir derfor 52 214,40 kroner, noe som danner grunnlag for å påstå at det eksisterer en bylønnspremie i Norge. Dette er de samme tallene som er presentert i gjennomgangen av deskriptiv statistikk.

Modell (1.2): Bylønnspremie i prosent

Som beskrevet i delkapittel 5.2 om modellene, modifiseres modell (1.1) fra en level-level modell til en log-level modell i (1.2). Løsningen av modellen blir:

$$\log_lønn = 10,76 + 0,1032 * \text{urbant}$$

Konstantleddet, β_0 , representerer logaritmen av lønn i rurale områder når dummyvariabelen for urbant er lik null, en verdi som i denne modellen er 10,76. Videre tilsier koeffisienten β_1 at urbane områder kan forventes å ha 10,32 prosentpoeng høyere lønn enn rurale områder. Bylønnspremien i Norge uttrykt i prosent kan derfor forventes å være over 10%. Den justerte determinasjonskoeffisienten i denne modellen er 0,1783. Dette tilsier at dummyvariabelen urbant ikke har særlig god forklaringskraft, og det behøves flere uavhengige variabler for å kunne fastslå de demografiske faktorene som påvirker lønnsgapet mellom urbane og rurale områder. Med hensikt å teste om geografisk område har en signifikant betydning for lønnsnivået i kommunen, testes dette i en hypotesetest med valgt signifikansnivå på 1 %:

$$H_0: \beta_1 = 0$$

$$H_A: \beta_1 \neq 0$$

$$TS = \frac{0,1032 - 0}{0,0114} = 9,053$$

$$t_{0,005}(355) = 2,576$$

Ved å lese av tabellen for t-distribusjon, kan kritisk verdi for 1% signifikansnivå og 354 frihetsgrader leses av som 2,576. Fra hypotesetesten kommer det frem at testobservasjonsverdien er 9,053, en verdi som er høyere enn den kritiske verdien. Ut fra dette kan det konkluderes med å forkaste H_0 , og at geografisk område har en signifikant betydning.

5.5 Modell (2): Gjennomsnittlig månedslønn med hensyn til utdanning og kjønn

Modell (2.1): Bylønnspremie med hensyn til utdanning

Modellen tar for seg hvordan utdanning, i tillegg til geografisk område, påvirker den gjennomsnittlige månedslønnen i en kommune. Den avhengige variabelen er logaritmen av gjennomsnittlig månedslønn også i denne modellen. Den første uavhengige variabelen er dummyvariabelen *urbant* som illustrerer hvor mye man kan forvente å øke i gjennomsnittlig månedslønn dersom kommunen klassifiseres som urban, alt annet likt. De resterende uavhengige variablene består av andel av befolkningen som har henholdsvis videregående skole, fagskole, kort høyere utdanning og lang høyere utdanning som høyeste fullførte utdanning. Disse variablene forteller hvor mye det kan forventes at et prosentpoengs endring av disse andelene øker eller reduserer den gjennomsnittlige månedslønnen, alt annet likt. Løsning av modellen:

$$\log_lønn = 10,63 + 0,0373 * \text{urbant} - 0,0006 * \text{andel_vgs} + 0,0211 * \text{andel_fagsk} + 0,0015 * \text{andel_høyereutd_kort} + 0,0087 * \text{andel_høyereutd_lang}$$

Konstantleddet i modellen indikerer at logaritmen av gjennomsnittlig månedslønn for en kommune er 10,63 dersom alle de andre uavhengige variabler hadde vært lik 0. Koeffisienten for *urbant* forteller at den gjennomsnittlige månedslønnen forventes å øke med 3,73 prosentpoeng i tilfeller hvor kommunen klassifiseres som urban. Dersom andelen av befolkningen som har videregående skole som høyeste fullførte utdanningsnivå øker med ett prosentpoeng, tilsier den tilhørende koeffisienten at den gjennomsnittlige månedslønnen forventes å reduseres med 0,06 prosentpoeng. Dersom andelen av befolkningen som har fagskole som høyeste fullførte utdanningsnivå øker med ett prosentpoeng, tilsier den tilhørende koeffisienten at den

gjennomsnittlige månedslønnen forventes å øke med 2,11 prosentpoeng. Denne tolkningen vil være lik også for kort og lang høyere utdanning, men den gjennomsnittlige månedslønnen forventes da å øke med henholdsvis 0,15 og 0,87 prosentpoeng. I denne modellen er den justerte R^2 på 0,4276. Denne verdien er betraktelig sterkere enn i modell (1.2). Med andre ord vil denne modellen, som i tillegg inkluderer utdanning, i større grad forklare lønnsgapet mellom urbane og rurale områder. For å vurdere signifikansen av utdannings effekt på gjennomsnittlig månedslønn, benyttes hypotesetesting med F-distribusjon med signifikansnivå på 1%.

$$H_0: \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$$

$$H_A: \text{Ikke } H_0$$

$$TS = \frac{(0,4286 - 0,1783)/4}{(1 - 0,4286)/(356 - 5)} = 27,261$$

$$F_{0,01}(4, 351) \approx F_{0,01}(4, 300) = 3,382$$

Forkastningsområdet består av verdier lik eller større enn 3,382. Siden testobservasjonsverdien på 27,261 ligger innenfor forkastningsområdet, støtter dette at H_0 kan forkastes, og det konkluderes derfor med at utdanning har en signifikant betydning for den gjennomsnittlige månedslønnen i en kommune. Det testes i tillegg for om geografisk område fremdeles har en signifikant betydning på gjennomsnittlig månedslønn:

$$H_0: \beta_1 = 0$$

$$H_A: \beta_1 \neq 0$$

$$TS = \frac{0,0372 - 0}{0,0118} = 3,153$$

$$t_{0,005}(355) \approx t_{0,005}(100) = 2,626$$

Forkastningsområdet i denne modellen blir dermed verdier mindre enn $-2,626$ og større enn $2,626$. Siden verdien til testobservasjonen ligger innenfor forkastningsområdet, forkastes H_0 , og det konkluderes med at effekten av at en kommune er urban eller rural fremdeles er av signifikans.

Modell (2.2): Bylønnspremie med hensyn til utdanning og kjønn

Modellen er en modifisert versjon av modell (2.1), ved at den inkluderer kjønn i tillegg til geografisk område og utdanning. På denne måten vil effekten av kjønn på gjennomsnittlig månedslønn bli presentert, gitt ulike utdanningsnivå. I likhet med modell (2.1) er den første uavhengige variabelen logaritmen av gjennomsnittlig månedslønn. Dummyvariabelen urbant er fremdeles inkludert da denne gir effekten både utdanning og alder har på gjennomsnittlig månedslønn når det skilles mellom urbane og rurale områder. De resterende uavhengige variablene inkluderer andel av befolkningen som har henholdsvis videregående skole, fagskole, kort høyere utdanning og lang høyere utdanning som høyeste fullførte utdanning, fordelt på kjønn. Løsning av modellen:

$$\begin{aligned} \log_l\ddot{o}nn = & 10,79 + 0,0367 * \text{urbant} - 0,0054 * \text{andel_vgs_kvinner} - 0,0094 * \\ & \text{andel_fagsk_kvinner} - 0,0039 * \text{andel_h\ddot{o}yereutd_kort_kvinner} - 0,0033 * \\ & \text{andel_h\ddot{o}yereutd_lang_kvinner} + 0,0028 * \text{andel_vgs_menn} + 0,0156 * \text{andel_fagskole_menn} + \\ & 0,0038 * \text{andel_h\ddot{o}yereutd_kort_menn} + 0,0113 * \text{andel_h\ddot{o}yereutd_lang_menn} \end{aligned}$$

I denne modellen er konstantleddet på 10,79. Verdien på konstantleddet antyder at logaritmen av gjennomsnittlig månedslønn for en kommune er 10,79 dersom alle de andre uavhengige variabler hadde vært lik 0. Modellen illustrerer tydelig hvordan kjønn i forhold til utdanning påvirker lønn. Alle de uavhengige variablene som inkluderer kvinner fører til en reduksjon i ulik grad i en kommunes gjennomsnittlige månedslønn, alt annet likt. For menn vil alle variablene derimot trekke mot en økning i en kommunes lønn. I denne modellen er den justerte R^2 betydelig høyere enn i de foregående modellene, med en verdi på 0,4881. Det vil med andre ord si at dersom kjønn inkluderes i modellen i tillegg til utdanning og urbant, vil forklaringskraften til modellen øke. For å teste om utdanning med hensyn til kjønn har en signifikant påvirkning på gjennomsnittlig månedslønn, utføres en hypotesetest med signifikansnivå på 1% også her:

$$H_0: \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = 0$$

$$H_A: \text{Ikke } H_0$$

$$TS = \frac{(0,4881 - 0,1783)/8}{(1 - 0,4881)/(356 - 9)} = 25,800$$

$$F_{0,01}(8, 347) \approx F_{0,01}(8, 300) = 2,571$$

Forkastningsområdet i dette tilfellet består av verdier lik eller større enn 2,571. Som følge av at testobservasjonsverdien på 25,800 ligger innenfor forkastningsområdet, støtter dette at H_0 kan forkastes, og det konkluderes derfor med at utdanning med hensyn til kjønn har en signifikant betydning for lønnen i en kommune. For å kunne konkludere om geografisk område fremdeles er av betydning for den gjennomsnittlige månedslønnen, testes også dette med et signifikansnivå på 1%:

$$H_0: \beta_1 = 0$$

$$H_A: \beta_1 \neq 0$$

$$TS = (0,0367 - 0) / 0,0115 = 3,191$$

$$t_{0,005}(355) \approx t_{0,005}(100) = 2,626$$

Forkastningsområdet i denne modellen blir dermed verdier mindre enn $-2,626$ og større enn $2,626$. Siden verdien til testobservasjonen er $3,191$ ligger det innenfor forkastningsområdet. Dermed forkastes H_0 , og det konkluderes med at effekten av at en kommune klassifiseres som urban eller rural fremdeles er signifikant.

5.6 Modell (3): Bylønnspremie med hensyn til alder

Alder blir hensyntatt i denne modellen, ved å sammenligne lønn i urbane og rurale områder fordelt på ulike aldersgrupper. Løsning av modellen:

$$\log_lønn = 11,02 + 0,0833 * \text{urbant} + 0,0029 * \text{andel_25til34år} - 0,0048 * \text{andel_35til44år} - 0,0057 * \text{andel_45til69år} - 0,0051 * \text{andel_70til79år} + 0,0022 * \text{andel_80ogeldre}$$

Konstantleddet i denne modellen er $11,02$, en betydelig høyere verdi enn for de foregående modellene. Resultatet kan med dette tolkes som at lønnen i rurale områder forventes å være høyere dersom alder og geografisk område utelukkende hensyntas, sammenlignet med når andre faktorer inkluderes. Videre observeres det at gjennomsnittlig månedslønn kan forventes å være $8,33$ prosentpoeng høyere om kommunen klassifiseres som urban. Dette er en økning fra modell (2.1) og modell (2.2), og en redusering fra modell (1.2), som indikerer at alder har mindre betydning for bylønnspremien enn utdanning og kjønn. De to sistnevnte faktorene kan med andre

ord forklare lønnsgapet mellom urbane og rurale områder i større grad. Videre fra modellen observeres det at aldersgruppene 25 til 34 år og de over 80 år, bidrar til en økning i en kommunes forventede gjennomsnittlige månedslønn. På en annen side bidrar alle aldersgrupper mellom 34 til 80 år til en redusering. Den justerte determinasjonskoeffisienten for denne modellen har en verdi på 0,2861, noe som tilsier at modellen ikke har spesielt god forklaringskraft. Likevel vil modell (3) anses som bedre forklarende for lønnsgapet enn modell (1.2), på bakgrunn av at denne verdien kun var på 0,1783 i modellen som utelukkende inkluderer geografisk område. Den justerte R^2 hadde en verdi på 0,4357 i modell (2.1) og 0,4881 i modell (2.2), hvilket indikerer at disse modellene har best forklaringskraft med hensyn til problemstillingen. Det utføres en hypotesetest for å teste signifikansen alder har på gjennomsnittlig månedslønn med signifikansnivå på 1%:

$$H_0: \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$$

$$H_A: \text{Ikke } H_0$$

$$TS = \frac{(0,2738 - 0,1783)/8}{(1 - 0,2738)/(356 - 5)} = 5,667$$

$$F_{0,01}(5, 351) \approx F_{0,01}(5, 300) = 3,079$$

Forkastningsområdet i dette tilfelle vil være for alle verdier lik eller større enn 3,079. På bakgrunn av at testobservasjonen ligger innenfor forkastningsområdet, kan det konkluderes med å forkaste H_0 , som med andre ord vil si at alder har en signifikant betydning for den gjennomsnittlige månedslønnen. For å undersøke om geografisk område fremdeles har en signifikant betydning på gjennomsnittlig månedslønn, utføres en t-test med 1% signifikansverdi:

$$H_0: \beta_1 = 0$$

$$H_A: \beta_1 \neq 0$$

$$TS = (0,0833 - 0) / 0,0115 = 7,243$$

$$t_{0,005}(355) \approx t_{0,005}(100) = 2,626$$

For denne hypotesetesten blir forkastningsområdet verdier mindre enn $-2,626$ og større enn $2,626$. Testobservasjonsverdien på 7,243 ligger innenfor dette området, og i likhet med de foregående modellene, har geografisk område fremdeles signifikant betydning.

5.7 Regresjonsresultat

Resultatet fra modell (1.2) fastslår at det eksisterer en bylønnspremie på over 10%, noe som ga grunnlag for å videre utforske hvorvidt denne blir påvirket av utdanning, kjønn og alder.

Funnene fra regresjonsanalysen blir presentert ved bruk av tabeller for en ryddig og lettlest oversikt. Tabell 4 illustrerer resultatene fra modell (2.1) og (2.2), mens Tabell 5 illustrerer resultatene fra modell (3).

	Effekt på gjennomsnittlig månedslønn		
	Kvinner	Menn	Total
Urban	-	-	3,72%
Andel med videregående skole	-0,54%	0,28%	-0,06%
Andel med fagskole	-0,94%	1,56%	2,11%
Andel meg kort høyere utdanning	-0,39%	0,38%	0,15%
Andel meg lang høyere utdanning	-0,33%	1,14%	0,87%

Tabell 4: Regresjonsresultat for Total er hentet fra modell 2.1. Resultater for Kvinner og Menn er hentet fra modell 2.2.

Ved å inkludere utdanning som forklaringsfaktor reduseres bylønnspremien til 3,72%, der andelen med fagskole som høyeste fullførte grad har størst effekt. Dersom utdanning fordelt på kjønn inkluderes som forklaringsfaktor, reduseres bylønnspremien til 3,67%. Det fremgår fra tabellen at det forventes en generell positiv effekt av andel menn på gjennomsnittlig månedslønn i en kommune. Motsatt er det forventet en generell negativ sammenheng mellom gjennomsnittlig månedslønn i en kommune og andel kvinner i kommunens befolkning. Disse trendene er henholdsvis positiv og negativ uavhengig av utdanningsnivå. En økning i andelen med fagskole er den uavhengige variabelen som forventes å gi størst positiv effekt på den lønnen, uavhengig av kjønn. For både kvinner og menn gir andel med fagskole som høyeste fullførte utdanning størst effekt på gjennomsnittslønnen i en kommune.

I en modell som utelukkende hensyntar alder og geografisk område vil bylønnspremien være på 8,22%. Det er andelen i befolkningen mellom 35-79 år som har størst effekt på gjennomsnittlig

månedslønn i en kommune. Effekten er negativ for denne aldersgruppen, mens trenden er positiv for befolkningen i aldersgrupper over og under. Dette illustreres i Tabell 5.

	Effekt på gjennomsnittlig månedslønn
Urban	8,22%
Andel 25 til 34 år	0,29%
Andel 35 til 44 år	-0,48%
Andel 45 til 69 år	-0,57%
Andel 70 til 79 år	-0,51%
Andel over 80 år	0,22%

Tabell 5: Regresjonsresultat for modell (3).

5.8 Robusthetsanalyse

Det finnes flere ulike måter å definere skillet mellom urbane og rurale områder på, som nevnt i kapittel 2.1. For å teste robustheten ved definisjonen som er tatt for seg i denne oppgaven, og følgelig resultatene av analysen, er det derfor valgt å utføre en robusthetsanalyse. Dette er interessant for å se om definisjonen av urbane og rurale områder er av betydning for resultatene ved en analyse av bylønnspremien i Norge. Det utføres robusthetsanalyse for modell (1.2), og det foretas to tester med et annet grunnlag for skillet mellom urbane og rurale kommuner enn i regresjonsanalysene.

I første del av robusthetsanalysen endres utvalget av urbane områder fra datasettet basert på sentralitetsindeksen. Hittil har kommuner med sentralitetsnivå 1 og 2 inngått i denne kategorien. Urbane områder består med dette kun av 25 kommuner. Det blir dermed vanskelig å unngå at man kun står igjen med de aller største byene, og områdene rundt disse, i denne sammenligningen. På bakgrunn av dette inkluderes de mellomsentrale kommunene med sentralitetsnivå 3 som urbane, for å se i hvilken grad dette påvirker resultatet. Dette kan være interessant å teste da det er rimelig å anta at byer som Ålesund, Tromsø og Kristiansand også for mange vil oppfattes som urbane, sammenlignet med områdene rundt.

Løsningen av modellen blir:

$$\log_l\ddot{o}nn = 10,75 + 0,0568 * \text{urbant_niv}\ddot{a}123$$

Konstantleddet i modellen er tilnærmet likt som i modell (1.2), på 10,75. Denne verdien indikerer at logaritmen av gjennomsnittlig månedslønn ville vært omtrent den samme som ved den tidligere definisjonen av urbane områder, dersom koeffisienten til den nye variabelen for urbant hadde vært lik 0. Koeffisienten til den nye uavhengige variabelen er derimot forventet å være 0,0568 i dette tilfellet, mot 0,1032 i modell (1.2). Med dette forstås at bylønnspremien ville vært forventet 4,64 prosentpoeng lavere enn ved opprinnelig definisjon. Observasjonen tyder på en forskjell i forventet bylønnspremie basert på hvilke områder man definerer som urbane.

Videre undersøkes endringen i bylønnspremien ved å definere et urbant område som en kommune med 150 000 innbyggere eller mer. Ved tidligere forskning har Carlsen et al. (2016) benyttet denne definisjonen. Denne robusthetstesten er relevant for å sammenligne med tidligere litteratur, og for å utforske oppgavens bidrag til litteraturen ved definisjon av urbane områder på en alternativ måte. Løsning for modellen:

$$\log_l\ddot{o}nn = 10,77 + 0,1713 * \text{urbant_150000}$$

I likhet med den tidligere robusthetstesten vil konstantleddet også i denne modellen være omtrent likt, uansett hvilken definisjon av urbane områder som brukes. På bakgrunn av dette kan det antas at gjennomsnittlig månedslønn for rurale kommuner ikke endres betraktelig ut fra hvilken definisjon som benyttes. Koeffisienten til den nye uavhengige variabelen er på 0,1713, som indikerer en bylønnspremie på 17,13% dersom denne definisjonen benyttes. Dette er en økning på 6,82 prosentpoeng fra den opprinnelige definisjonen, og en økning på 11,45 prosentpoeng fra analysen hvor kommuner med sentralitetsnivå 3 ble inkludert som urbane.

Oppsummert påvirkes bylønnspremien betraktelig ved endring av definisjonen av urbane områder, samtidig som endringen for gjennomsnittlig månedslønn i rurale kommuner ikke er bemerkelsesverdig. Videre øker bylønnspremien desto høyere krav som stilles for å definere en kommune som urban.

Valget om å kategorisere urbane kommuner etter sentralitetsnivå 1 og 2 antas også etter robusthetsanalysen å være hensiktsmessig for å gi svar på problemstillingen. Dette med tanke på hvilke demografiske faktorer som legges til grunn for å analysere gapet og ønsket om å finne de ekstreme verdiene. I tillegg har kun et fåtall av norske kommuner minst 150 000 innbyggere, noe som indikerer at denne definisjonen av urbane områder ikke vil være passende for Norges kommuner. Funnet fra robusthetsanalysen tyder følgelig på at definisjonen i denne oppgaven kan være et tilskudd til tidligere litteratur. Resultatene fra robusthetstesten som definerer kommuner med minst 150 000 innbyggere som urbane, viser at bylønnspremien øker betraktelig sammenlignet med tidligere definisjon. Motsatt reduseres bylønnspremien omtrent like mye ved å inkludere sentralitetsnivå 3. Dette kan tyde på at definisjonen som er valgt i denne oppgaven, kan anses som en effektiv definisjon for beregning av bylønnspremien i Norge. Robusthetstestene viser at resultatene som fremkommer av regresjonsanalysen kun vil være gjeldende ved urbane områder definert som kommuner med sentralitetsnivå 1 og 2.

6. Diskusjon

I dette delkapittelet skal regresjonsresultatene tolkes. Disse tolkningene danner grunnlag for anbefalinger til videre forskning. Til slutt blir eventuelle feilkilder presentert og diskutert.

6.1 Tolking av regresjonsresultater

Etter beregning av regresjonsmodellene i oppgaven, kan det konkluderes med en lønnsforskjell på 10,32% mellom urbane og rurale områder i Norge. Funnet indikerer at det eksisterer en bylønnspremie, i likhet med resultater fra tidligere litteratur presentert i oppgaven. Denne likheten kan anses å styrke troverdigheten i oppgavens empiriske funn.

Effekten av utdanning og kjønn

Det illustreres i modell (2.2) at en kommunes utdanningsdistribusjon fordelt på kjønn er av stor statistisk signifikans. Dette styrkes av at modellen har høyest determinasjonskoeffisient av samtlige modeller. Til tross for den lavere verdien for R^2 i de øvrige modellene, vil disse likevel gi resultater av statistisk signifikans. I artiklene fra litteraturgjennomgangen ble kjønn og utdanning vektlagt som faktorer av betydning for lønn. Dette støttes av resultatene fra modell (2.2), som gjør at funnene i denne analysen støttes opp under tidligere empiri. Fra den deskriptive statistikken kommer det frem at en større andel av befolkningen i urbane områder har høyere utdanning. Dette kan tenkes å være en effekt som bidrar til et lønnsgap mellom urbane og rurale områder, noe som også vises i regresjonsresultatene. Den samme forklaringen for bylønnspremie omtales i tidligere litteratur om sorteringstendens (de la Roca & Puga, 2017), samt i litteratur om assortativ matching (Dauth et al. 2022; Leknes et al. 2022). Alle artiklene tar for seg trenden om at individer med høy produktivitet og utdanning trekker til byene, hvilket resulterer i en bylønnspremie.

Den negative koeffisienten til andel videregående skole indikerer at den gjennomsnittlige månedslønnen i kommunen reduseres dersom andelen av befolkningen som velger å ikke ta høyere utdanning øker. Avkastning på utdanning er høyere for utdannede på masternivå enn på bachelornivå, noe vi kan se i modell (2.1). Den betydelig høyere andelen med fullført

bachelorgrad enn mastergrad som observeres i den deskriptive statistikken, kan være et argument for denne sammenhengen. Koeffisienten tilhørende andelen med fagskole som høyeste utdanningsnivå utmerker seg med å være relativt høyere enn koeffisientene for andelen av de andre utdanningene. Dette kan ha en praktisk forklaring. Andelen mennesker i en befolkning som har fagskole som høyeste fullførte utdanning er relativt lav, uavhengig av om kommunen klassifiseres som urban eller rural. En prosentvis økning i andelen med fagskole vil derfor føre til en høyere marginaleffekt, enn en prosentvis endring i andelen med en annen utdanning.

Analysens deskriptive statistikk tilsier videre at bylønnspremien er omtrent dobbelt så høy for menn som for kvinner. Dette i motsetning til forskningen fra Leknes et al. (2022), som heller kommer frem til at det er kvinner som drar størst nytte av å bo i byen. Det kan tenkes at disse forskjellene faller på bakgrunn av ulik metode, med eksempelvis ulike forklarende variabler inkludert i modellene, ulike datasett med ulike tilnæringer til definisjon av urbane områder og ulikt tidsperspektiv. Videre observeres det fra oppgavens analyse at gjennomsnittlig månedslønn i en kommune påvirkes negativt ved en økning i andel kvinner, uavhengig av utdanningsnivå. På bakgrunn av disse funnene kan det konkluderes med at det eksisterer et lønnsgap mellom kvinner og menn uavhengig av geografisk område. Tolkningen av dette er at kjønn vil påvirke lønnsforskjeller i urbane og rurale områder i form av andel kvinner og menn i kommunen. En kommune bestående av en høy andel kvinner vil følgelig gi lavere gjennomsnittlig månedslønn i kommunen, sammenlignet med en kommune hvor den samme andelen heller er menn. Denne sammenhengen gjelder både for urbane og rurale kommuner.

Effekten av alder

Alder er en ytterligere demografisk faktor som kan bidra til forklaringen av lønnsforskjeller mellom urbane og rurale områder. Befolkningen i urbane områder kan tenkes å i større grad bestå av yngre mennesker, blant annet på bakgrunn av at mange yngre mennesker flytter til mer sentrale områder for å studere. Dette er en effekt som kan antas å fanges opp i modell (3). Resultatene fra denne modellen viser at alder har en relativt liten effekt på den gjennomsnittlige månedslønnen, sammenlignet med de andre uavhengige variablene. En overraskende observasjon fra resultatene er at det er de yngste og eldste individene som har en positiv effekt på

lønnen. På en annen side er det relevant å nevne at alder kun er en kontroll for å skille ut de ulike effektene aldersdistribusjon har på lønn i denne modellen. Hovedmålet med å beregne modellen vil derfor være å undersøke hvordan verdien for koeffisienten urbant endres ved å skille ut effekten alder forventes å ha på gjennomsnittlig månedslønn. Dummyvariabelen for urbant er 0,0833, noe som antyder at koeffisienten fortsatt har en relativt høy verdi i forhold til de øvrige modellene dersom man skiller ut alders isolerte effekt. Dette sett i sammenheng med modell (1.2), hvor koeffisienten for urbant i SLR-modellen var 0,1032. Med andre ord kan den isolerte effekten av alder på lønn tolkes som å ikke ha like stor effekt som utdanning og kjønn, og at en mulig forklaring kan baseres på ulik aldersdistribusjon i urbane og rurale områder.

Regresjonsresultatene til modell (3) tilsier motsatte effekter enn hva som fremkommer fra tidligere litteratur. Ifølge Dauth et al. (2022) er sammenhengen mellom bylønnspremie og alder utelukkende positiv. I likhet med forklaringene gitt ved sammenligning av bylønnspremien og kjønn, er det rimelig å anta at denne forskjellen mellom litteraturen og resultatene fra denne modellen delvis kan forklares på bakgrunn av ulik metode.

6.2 Oppgavens anbefaling til videre forskning

Tolkningen av koeffisientene og deres varierende effekt på gjennomsnittlig månedslønn i urbane og rurale områder, kan danne et grunnlag for anbefaling til videre forskning. En naturlig utvidelse av analysen kan være å inkludere ytterligere uavhengige variabler. Dette kan også sees i sammenheng med determinasjonskoeffisienten. På bakgrunn av at forklaringskraften til en modell ligger mellom 0 og 1, kan resultatene i oppgavens modeller anses å gi begrenset forklaring for bylønnspremien, med høyeste verdi på 0,4881. Verdien indikerer at det anses som nyttig å inkludere flere uavhengige variabler for å gi en styrket forklaring. Et skille mellom offentlig og privat sektor er et eksempel på ytterligere forklaringsvariabler. I tillegg er erfaring en variabel som kunne tenkes å være av betydning for resultatene, slik som forklart i delkapittel 3.4 om styrker og svakheter ved datagrunnlaget.

6.3 Feilkilder

Tallene for gjennomsnittslønn er i datasettene oppgitt som nominelle, et utgangspunkt som kan legge grunnlag for et annet resultat enn om man tar hensyn til forskjeller i kjøpekraft. Forklaringen på dette er høyere kostnader i urbane områder for både arbeidsgiver og arbeidstaker, sammenlignet med i rurale områder, som omtalt i delkapittel 2.1. Selv om analysen konstaterer at gjennomsnittslønnen er høyere i urbane områder, er det likevel ikke gitt at kjøpekraften følgelig er høyere, siden lønnen i dette tilfellet ikke er kjøpekraftjustert. Feilkilden er ikke hensyntatt i oppgaven, da problemstillingen ikke omfatter kjøpekraft, men heller tar for seg nominell gjennomsnittslønn.

I noen av kommunene er det ikke registrerte tall for alle de ulike variablene. Denne svakheten ved datasettet kan føre til feil i analysen ved at MLR. 2 muligens ikke holder. Det er likevel argumentert for at feilkilden er av minimal betydning i kapittel 4.3, hvilket gir grunnlag for at antakelsen om tilfeldig utvalg holder.

Ved at datasettet kun inneholder et øyeblikksbilde for 2022, gir dette rom for mulige feilkilder ved at det kan inneholde variabler som ikke representerer arbeidsmarkedet i dagens samfunn (2024). Det mangler også flere datapunkter som underbygger størrelsene i gitte kommuner. Samtidig anses datagrunnlaget såpass omfattende, at dette ikke skal medføre betydelig forvreining av funn. Det er heller ikke ønsket å belyse bylønnspremien over tid i problemstillingen, men vise et bilde på nåsituasjonen. Imidlertid kan det antas at lønnsbilde i 2022 var preget av koronapandemien. Dette kan ha betydning for resultatene i analysen, ved at flere effekter av pandemien vil antas å henge igjen i 2022.

7. Konklusjon

I denne oppgaven er det utført en empirisk analyse for å besvare problemstillingen om hvordan ulike demografiske faktorer påvirker lønnsforskjeller i urbane og rurale områder. De demografiske faktorene analysen har valgt å inkludere er geografisk område, utdanning, kjønn og alder. Hovedfunnene fra studien viser at alle de ovennevnte demografiske faktorene har en signifikant betydning for lønnsforskjeller i urbane og rurale områder. I likhet med tidligere litteratur kommer det frem fra samtlige modeller at geografisk område er av signifikans, uavhengig av hvilke øvrige faktorer som inkluderes i regresjonene, alt annet likt. Det kan på generelt grunnlag konkluderes med at det finnes en bylønnspremie i Norge. Blant øvrige uavhengige variabler er det utdanning som har størst effekt på lønn. Videre har menn en høyere bylønnspremie enn kvinner, og kvinner påvirker den gjennomsnittlige månedslønnen i en kommune negativt uavhengig av utdanningsnivå. Overraskende funn fra analysen er hvor liten effekt alder har på lønnsgapet, og hvilke aldergrupper som påvirker lønn i positiv og negativ retning.

I likhet med tidligere litteratur konstaterer oppgavens empiri at det eksisterer en bylønnspremie. En ytterligere likhet er utdanning som en driver til lønnsforskjeller i urbane og rurale områder. Det er flere høyt utdannede arbeidstakere i urbane områder, hvilket bidrar til at utdanning har betydelig effekt på bylønnspremien. Dette underbygges av tidligere empiri om sorteringstendens og assortativ matching. Analysen av kjønn og alder avviker derimot fra tidligere forskning.

Opgavens bidrag til litteraturen kommer til uttrykk ved definisjon av urbane og rurale områder som er basert på sentralitetsindeksen til SSB. Videre inkluderer analysen offentlig sektor i analysen, noe som anses hensiktsmessig da det ønskes en nyansert fremstilling av det totale arbeidsmarkedet, i tillegg til at det er en større andel av offentlige arbeidsplasser i urbane områder i Norge.

En stor del av svakhetene ved oppgaven omhandler begrensninger ved datasettet. Manglende tall for enkelte kommuner kan være en delvis forklaring for eventuelle feil i resultatene, og dermed også en del av forklaringen av hvorfor resultatene i denne analysen på noen punkter avviker fra

tidligere litteratur. Datasettets få kategorier av variabler kan betraktes som en ytterligere svakhet. Som konsekvens av dette fremkommer det at determinasjonskoeffisienten er relativt lite forklarende for bylønnspremien. Ved videre forskning på emnet kan det derfor tenkes å være nyttig å inkludere flere uavhengige variabler for å gi en mer helhetlig forklaring av lønnsforskjeller i urbane og rurale områder i Norge.

8. Referanseliste

Anderson, D. R., Sweeny, D. J., Williams, T. A., Camm, J. D., Cochran, J. J., Fry, M. J. & Ohlmann, J. W. (2020). *Essentials of Modern Business Statistics*. Cengage.

Carlsen, F., Rattsø, J., & Stokke, H. E. (2016). Education, experience, and urban wage premium. *Regional Science and Urban Economics*, 60, 39-49.

<https://doi.org/10.1016/j.regsciurbeco.2016.06.006>

Dauth, W., Findeisen, S., Moretti, E., & Suedekum, J. (2022). Matching in cities. *Journal of the European Economic Association*, 20(4), 1478-1521.

<https://doi.org/10.1093/jeea/jvac004>

De la Roca, J., & Puga, D. (2017). Learning by working in big cities. *The Review of Economic Studies*, 84(1), 106–142. <https://doi.org/10.1093/restud/rdw031>

Leknes, S., Rattsø, J. & Stokke, H. E. (2022). Assortative labor matching, city size, and the education level of workers. *Regional Science and Urban Economics*, 96(103806), 1-18.

<https://doi.org/10.1016/j.regsciurbeco.2022.103806>

Høydahl, E. (2020). *Sentralitetsindeksen. Oppdatering med 2020-kommuner*. Statistisk sentralbyrå. Hentet 22.04.24. <https://ssb.brage.unit.no/ssb-xmlui/handle/11250/2649672>

Høydahl, E. (2023, 20. juni). *Sentralitetsindeksen*. Statistisk sentralbyrå. Hentet 10.05.24. <https://www.ssb.no/befolkning/folketall/artikler/sentralitetsindeksen>

Statistisk sentralbyrå. (2023). *09429: Utdanningsnivå, etter kommune og kjønn (K) 1970 – 2022* [Statistikk]. Hentet 16.04.2024. <https://www.ssb.no/statbank/table/09429>

Statistisk sentralbyrå. (2024). 07459: *Alders- og kjønnsfordeling i kommuner, fylker og hele landets befolkning (K) 1986 – 2024*. Hentet 10.05.2024.

<https://www.ssb.no/statbank/table/07459/>

Statistisk sentralbyrå. (2024). 12852: *Kommunefordelt månedslønn, etter bosted, arbeidssted, alder og kjønn (K) 2015 – 2023* [Statistikk]. Hentet 16.04.2024.

<https://www.ssb.no/statbank/table/12852>

Sucarrat, G. (2017). *Metode og økonometri: en moderne innføring*. (2. utg.) Fagbokforlaget.

Wooldridge, J. M. (2020). *Introductory Econometrics: A Modern Approach*. (7. utg.). Cengage.

9. Vedlegg

Vedlegg 1: Forklaring av variabler og koeffisienter i regresjonsmodellene

Dette vedlegget inneholder forklaringer av alle oppgavens variabler og koeffisienter. Det er verdt å nevne at flere av de uavhengige variablene oppgir andel av befolkningen som en verdi mellom 0 og 100, og kan derfor forstås som en prosent av befolkningen. Med unntak av modell (1.1) er samtlige modeller en log-level modell, noe som har betydning for hvordan man tolker koeffisientene. Det er derfor viktig å ikke tolke de uavhengige variablene som en log-form, da dette ville ha gitt implikasjoner om at modellene er en log-log modell, noe som ville ha gitt feil tolkninger av koeffisientene.

Modell (1.1):

Avhengig variabel:

gj_lønn: gjennomsnittlig månedslønn uttrykt i kroner

Uavhengig variabel:

urbant: dummyvariabel for om en kommune anses som urban eller ikke. Variabelen har verdien 1 om kommunen klassifiseres som urban, og 0 om den klassifiseres som rural.

Koeffisienter:

b_0 : konstantleddet som forklarer hva den avhengige variabelen, gjennomsnittlig månedslønn i kroner, forventes å bli gitt at den uavhengige variabelen *urbant* er lik null.

b_1 : forteller hvor mange prosentpoeng logaritmen til gjennomsnittlig månedslønn forventes å endres ved at kommunen klassifiseres som urban.

Modell (1.2):

Avhengig variabel:

log_lønn: logaritmen av gjennomsnittlig månedslønn

Uavhengig variabel:

urbant: dummyvariabel for om en kommune anses som urban eller ikke. Variabelen har verdien 1 om kommunen klassifiseres som urban, og 0 om den klassifiseres som rural.

Koeffisienter:

b_0 : konstantleddet som forklarer hva den avhengige variabelen, logaritmen av gjennomsnittlig månedslønn, vil bli gitt at den uavhengige variabelen *urbant* er lik null.

b_1 : forteller hvor mange prosentpoeng logaritmen til gjennomsnittlig månedslønn forventes å endres ved at kommunen klassifiseres som urban.

Modell (2.1):

Avhengig variabel:

log_lønn: logaritmen av gjennomsnittlig månedslønn

Uavhengige variabler:

urbant: dummyvariabel for om en kommune anses som urban eller ikke. Variabelen har verdien 1 om kommunen klassifiseres som urban, og 0 om den klassifiseres som rural.

andel_vgs: Andelen av en kommunes befolkning som har videregående skole som høyeste grad av fullført grad av utdanning. Oppgitt som en verdi mellom 0 og 100.

andel_fagsk: Andelen av en kommunes befolkning som har fagskole som høyeste grad av fullført grad av utdanning. Oppgitt som en verdi mellom 0 og 100.

andel_høyereutd_kort: Andelen av en kommunes befolkning som har under 4 års høyere utdanning fra universitet eller høyskole som høyeste grad av fullført grad av utdanning. Oppgitt som en verdi mellom 0 og 100.

andel_høyereutd_lang: Andelen av en kommunes befolkning som har over 4 års lang høyere utdanning fra universitet eller høyskole som høyeste grad av fullført grad av utdanning. Oppgitt som en verdi mellom 0 og 100.

Koeffisienter:

b_0 : konstantleddet som forklarer hva den avhengige variabelen, logaritmen av gjennomsnittlig månedslønn, vil bli gitt at alle uavhengige variabler er lik null.

b_1 : forteller hvor mange prosentpoeng logaritmen til gjennomsnittlig månedslønn forventes å endres ved at kommunen klassifiseres som urban.

b_2 : forteller hvor mange prosentpoeng logaritmen til gjennomsnittlig månedslønn forventes å endres ved at kommunens andel av befolkningen som har videregående skole som høyeste fullførte utdanning øker med ett prosentpoeng.

b_3 : forteller hvor mange prosentpoeng logaritmen til gjennomsnittlig månedslønn forventes å endres, ved at kommunens andel av befolkningen som har fagskole som høyeste fullførte utdanning øker med ett prosentpoeng.

b_4 : forteller hvor mange prosentpoeng logaritmen til gjennomsnittlig månedslønn forventes å endres, ved at kommunens andel av befolkningen som har kort høyere utdanning som høyeste fullførte utdanning øker med ett prosentpoeng.

b_5 : forteller hvor mange prosentpoeng logaritmen til gjennomsnittlig månedslønn forventes å endres, ved at kommunens andel av befolkningen som har lang høyere utdanning som høyeste fullførte utdanning øker med ett prosentpoeng.

Modell (2.2):

Avhengig variabel:

log_lønn: logaritmen av gjennomsnittlig månedslønn

Uavhengige variabler:

urbant: dummyvariabel for om en kommune anses som urban eller ikke. Variabelen har verdien 1 om kommunen klassifiseres som urban, og 0 om den klassifiseres som rural.

andel_vgs_menn: Andelen av en kommunes mannlige befolkning som har videregående skole som høyeste grad av fullført grad av utdanning. Oppgitt som en verdi mellom 0 og 100.

andel_fagsk_menn: Andelen av en kommunes mannlige befolkning som har fagskole som høyeste grad av fullført grad av utdanning. Oppgitt som en verdi mellom 0 og 100.

andel_høyereutd_kort_menn: Andelen av en kommunes mannlige befolkning som har en under 4 års lang høyere utdanning fra universitet eller høyskole som høyeste grad av fullført grad av utdanning. Oppgitt som en verdi mellom 0 og 100.

andel_høyereutd_lang_menn: Andelen av en kommunes mannlige befolkning som har en over 4 års lang høyere utdanning fra universitet eller høyskole som høyeste grad av fullført grad av utdanning. Oppgitt som en verdi mellom 0 og 100.

andel_vgs_kvinner: Andelen av en kommunes kvinnelige befolkning som har videregående skole som høyeste grad av fullført grad av utdanning. Oppgitt som en verdi mellom 0 og 100.

andel_fagsk_kvinner: Andelen av en kommunes kvinnelige befolkning som har fagskole som høyeste grad av fullført grad av utdanning. Oppgitt som en verdi mellom 0 og 100.

andel_høyereutd_kort_kvinner: Andelen av en kommunes kvinnelige befolkning som har en under 4 års lang høyere utdanning fra universitet eller høyskole som høyeste grad av fullført grad av utdanning. Oppgitt som en verdi mellom 0 og 100.

andel_høyereutd_lang_kvinner: Andelen av en kommunes kvinnelige befolkning som har en over 4 års lang høyere utdanning fra universitet eller høyskole som høyeste grad av fullført grad av utdanning. Oppgitt som en verdi mellom 0 og 100.

Koeffisienter:

b_0 : konstantleddet som forklarer hva den avhengige variabelen, logaritmen av gjennomsnittlig månedslønn, vil bli gitt at alle uavhengige variabler er lik null.

b_1 : forteller hvor mange prosentpoeng logaritmen til gjennomsnittlig månedslønn forventes å endres ved at kommunen klassifiseres som urban.

b_2 : forteller hvor mange prosentpoeng logaritmen til gjennomsnittlig månedslønn forventes å endres, ved at kommunens andel av den kvinnelige befolkningen som har videregående skole som høyeste fullførte utdanning øker med ett prosentpoeng.

b_3 : forteller hvor mange prosentpoeng logaritmen til gjennomsnittlig månedslønn forventes å endres, ved at kommunens andel av den kvinnelige befolkningen som har fagskole som høyeste fullførte utdanning øker med ett prosentpoeng.

b₄: forteller hvor mange prosentpoeng logaritmen til gjennomsnittlig månedslønn forventes å endres, ved at kommunens andel av den kvinnelige befolkningen som har kort høyere utdanning som høyeste fullførte utdanning øker med ett prosentpoeng.

b₅: forteller hvor mange prosentpoeng logaritmen til gjennomsnittlig månedslønn forventes å endres, ved at kommunens andel av den kvinnelige befolkningen som har lang høyere utdanning som høyeste fullførte utdanning øker med ett prosentpoeng

b₆: forteller hvor mange prosentpoeng logaritmen til gjennomsnittlig månedslønn forventes å endres, ved at kommunens andel av den mannlige befolkningen som har videregående skole som høyeste fullførte utdanning øker med ett prosentpoeng.

b₇: forteller hvor mange prosentpoeng logaritmen til gjennomsnittlig månedslønn forventes å endres, ved at kommunens andel av den mannlige befolkningen som har fagskole som høyeste fullførte utdanning øker med ett prosentpoeng.

b₈: forteller hvor mange prosentpoeng logaritmen til gjennomsnittlig månedslønn forventes å endres, ved at kommunens andel av den mannlige befolkningen som har kort høyere utdanning som høyeste fullførte utdanning øker med ett prosentpoeng.

b₉: forteller hvor mange prosentpoeng logaritmen til gjennomsnittlig månedslønn forventes å endres, ved at kommunens andel av den mannlige befolkningen som har lang høyere utdanning som høyeste fullførte utdanning øker med ett prosentpoeng

Modell (3):

Avhengig variabel:

log_lønn: logaritmen av gjennomsnittlig månedslønn

Uavhengige variabler:

urbant: dummyvariabel for om en kommune anses som urban eller ikke. Variabelen har verdien 1 om kommunen klassifiseres som urban, og 0 om den klassifiseres som rural.

andel_25til34år: En kommunes andel av befolkningen som er mellom 25 og 34 år.

Opgitt som en verdi mellom 0 og 100.

andel_35til44år: En kommunes andel av befolkningen som er mellom 35 og 44 år.

Opgitt som en verdi mellom 0 og 100.

andel_45til69år: En kommunes andel av befolkningen som er mellom 45 og 69 år. Oppgitt som en verdi mellom 0 og 100.

andel_70til79år: En kommunes andel av befolkningen som er mellom 70 og 79 år. Oppgitt som en verdi mellom 0 og 100.

andel_80ogeldre: En kommunes andel av befolkningen som er 80 år eller eldre. Oppgitt som en verdi mellom 0 og 100.

Koeffisienter:

b_0 : konstantleddet som forklarer hva den avhengige variabelen, logaritmen av gjennomsnittlig månedslønn, vil bli gitt at alle uavhengige variabler er lik null.

b_1 : forteller hvor mange prosentpoeng logaritmen til gjennomsnittlig månedslønn forventes å endres ved at kommunen klassifiseres som urban.

b_2 : forteller hvor mange prosentpoeng logaritmen til gjennomsnittlig månedslønn forventes å endres, ved at kommunens andel av befolkningen med alder mellom 25 og 34 år øker med ett prosentpoeng

b_3 : forteller hvor mange prosentpoeng logaritmen til gjennomsnittlig månedslønn forventes å endres, ved at kommunens andel av befolkningen med alder mellom 35 og 44 år øker med ett prosentpoeng

b_4 : forteller hvor mange prosentpoeng logaritmen til gjennomsnittlig månedslønn forventes å endres, ved at kommunens andel av befolkningen med alder mellom 45 og 69 år øker med et prosentpoeng

b_5 : forteller hvor mange prosentpoeng logaritmen til gjennomsnittlig månedslønn forventes å endres, ved at kommunens andel av befolkningen med alder mellom 70 og 79 år øker med ett prosentpoeng

b_6 : forteller hvor mange prosentpoeng logaritmen til gjennomsnittlig månedslønn forventes å endres, ved at kommunens andel av befolkningen som er 80 år eller eldre øker med ett prosentpoeng

Vedlegg 2: Sammenstilling av regresjonsresultater

VARIABLES	(1.1) gjlønn	(1.2) log_lønn	(2.1) log_lønn	(2.2) log_lønn	(3) log_lønn
urbant	5,060*** (563.4)	0.100*** (0.0114)	0.0372*** (0.0118)	0.0367*** (0.0117)	0.0833*** (0.0115)
andel_vgs_kvinner				-0.00542*** (0.00131)	
andel_fagsk_kvinner				-0.00942** (0.00396)	
andel_høyereutd_kort_kvinner				-0.00388*** (0.00118)	
andel_høyereutd_lang_kvinner				-0.00331 (0.00222)	
andel_vgs_menn				0.00282*** (0.000900)	
andel_fagsk_menn				0.0156*** (0.00170)	
andel_høyereutd_kort_menn				0.00378** (0.00150)	
andel_høyereutd_lang_menn				0.0113*** (0.00222)	
andel_vgs			-0.000606 (0.000713)		
andel_fagsk			0.0211*** (0.00300)		
andel_høyereutd_kort			0.00150 (0.00119)		
andel_høyereutd_lang			0.00874*** (0.00172)		
andel_25til34år					0.00292 (0.00217)
andel_35til44år					-0.00475 (0.00310)
andel_45til69år					-0.00574*** (0.00166)
andel_70til79år					-0.00511** (0.00244)
andel_80ogeldre					0.00220 (0.00329)
Constant	47,155*** (149.3)	10.76*** (0.00301)	10.63*** (0.0395)	10.79*** (0.0464)	11.02*** (0.0996)
Observations	356	356	356	356	356
R-squared	0.186	0.181	0.436	0.501	0.286

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Tabell 6: Tabellarisk utforming av modellenes regresjonsresultater

