

# **Assignment 4, MSB104-gruppe 6, 2025**

Karin Liang

## **Kort sammendrag**

Denne oppgaven undersøker regionale forskjeller i økonomisk utvikling og inntektsulikhet i Belgia, Nederland, Bulgaria og Norge ved hjelp av Eurostat-data. Analysene kombinerer tre tilnæringer: deskriptive analyser av BNP per innbygger og Gini, tverrsnittsregresjoner basert på 2017-data, samt alternative funksjonelle former og panelmodeller for perioden 2008 til 2022. Formålet er å vurdere hvordan økonomisk utvikling, utdanningsnivå, transportinfrastruktur og demografisk struktur henger sammen med regionale ulikhetsnivåer.

Resultatene viser at utdanningsnivå og veitethet er de mest konsistente forklaringsfaktorene for variasjon i inntektsulikhet, mens økonomisk vekst har svak og ustabil betydning på tvers av modellene. Panelestimeringene viser svært begrenset tidsvariasjon i utdanningsnivå og demografisk struktur, noe som gjør det vanskelig å identifisere tydelige tidsdynamiske effekter i regional ulikhet. Analysene gir et oversiktssbilde av hvordan et lite sett av variabler henger sammen med regional inntektsulikhet, og kan fungere som et utgangspunkt for videre empiriske analyser og vurderinger innen regionalpolitikken.

## **Innledning**

### **Bakgrunn**

Det finnes fortsatt tydelige forskjeller i økonomisk kapasitet mellom europeiske land og mellom regioner innen hvert land. Basert på BNP per innbygger data hentet fra Eurostat kan man se at både på nasjonalt og regionalt nivå viser økonomisk utvikling betydelig variasjon, både i langsiktige trender og i tverrsnittssammenligninger. Disse forskjellene gjenspeiler ikke bare variasjoner i regional produksjonskapasitet og økonomisk struktur, men er også nært knyttet til flere nøkkelfaktorer, blant annet utdanningsnivå (høy utdanning), transportinfrastruktur (veitethet), demografisk struktur (andel av befolkningen som er 65 år eller eldre), og nivået på eller endringen i BNP per innbygger.

### **Mål**

Denne oppgaven integrerer analysene fra oppgave 1, 2 og 3 for systematisk å undersøke regionale forskjeller i økonomisk utvikling og forholdet mellom regional utvikling og inntektsulikhet i Belgia (BE), Nederland (NL), Bulgaria (BG) og Norge (NO). Oppgavene fokuserer på sentrale faktorer som økonomisk utvikling, utdanningsnivå, transportinfrastruktur og demografisk struktur, og analyserer hvordan disse faktorene henger sammen med regional inntektsulikhet. Videre vurderes hvor robuste disse sammenhengene er på tvers av ulike datastrukturer og metodiske rammeverk.

Hovedmålet i oppgave 1 er å benytte BNP og befolkningsdata fra Eurostat til å konstruere BNP per innbygger (BNPC) og Gini koeffisienten på NUTS2 nivå for perioden 2000 til 2023 for de fire landene. Dette brukes til å sammenligne nivået på regional økonomisk utvikling og

de langsiktige utviklingstrendene. Gjennom en beskrivende analyse presenteres utviklingsmønstre for regionene, noe som gir et viktig bakteppe for videre analyser av regionale forskjeller og inntektsulikhet.

Oppgave 2 undersøker tverrsnittssammenhengen mellom regional utvikling og inntektsulikhet. Basert på tverrsnittsdata for 2017 estimeres en multiple lineær regresjonsmodell for å vurdere om økonomisk utvikling (endring i BNP per innbygger), transportinfrastruktur (veitetthet), utdanningsnivå (høy utdanning) og demografisk struktur (andel 65 år eller eldre) er systematisk forbundet med regionale nivåer av Gini koeffisienten.

Oppgave 3 utvider analysen metodisk ved å benytte alternative funksjonelle former, som log lineær modell, kvadratisk modell og kubisk modell, samt panelmodeller med faste effekter. Hensikten er å undersøke om de samme variablene viser robuste sammenhenger under ulike modellspecifikasjoner og i tidsdimensjonen. Paneldatastrukturen gjør det mulig å kontrollere for uobserverbare, tidsinvariante regionale egenskaper, noe som gir et mer helhetlig grunnlag for å analysere de dynamiske forholdene mellom regional utvikling og inntektsulikhet.

## Betydning

Forholdet mellom regional utvikling og inntektsulikhet er viktig å studere fordi økonomiske forskjeller mellom regioner påvirker inntektsnivået, levekårene og de langsiktige mulighetene for befolkningen. Å identifisere hvilke faktorer som henger sammen med ulikhet, som BNP per innbygger, utdanningsnivå, veitetthet eller demografisk struktur, gir grunnlag for mer treffsikre regionale politiske tiltak.

Ved å kombinere langtidsdata for BNP per innbygger fra oppgave 1, tverrsnittsresultater fra oppgave 2 og alternative funksjonelle former og faste effektsmodeller fra oppgave 3, undersøker disse oppgavene om de viktigste faktorene bak regional inntektsulikhet er stabile og pålitelige. Denne helhetlige tilnærmingen gir en bredere forståelse av forholdet mellom regional utvikling og inntektsulikhet, og kan bidra med et empirisk grunnlag for videre politiske vurderinger.

## Litteraturgjennomgang

### Tidligere arbeid

Denne delen gir en oversikt over relevant litteratur fra oppgave 1–3 og danner et faglig grunnlag for å forstå sammenhenger mellom regional utvikling, utdanning, infrastruktur, demografi og inntektsulikhet.

Lessmann & Seidel (2017) utnytter satellittdata for nattlys fra 180 land og 3166 regioner, og finner at om lag 67–70 % av landene viser sigma-konvergens, der inntektsforskjellene mellom regioner har blitt mindre over tid. Artikkelen avdekker også en N-formet sammenheng mellom økonomisk utvikling og regional ulikhet: ulikheten kan først øke, deretter avta, og igjen stige i høyinntektsfaser. Videre peker artikkelen på at utvikling, mobilitet, åpenhet, ressurser, institusjoner, overføringer og utdanning, og etnisitet utgjør syv grupper av faktorer som kan påvirke determinanter for regional ulikhet. Denne artikkelen gir et teoretisk grunnlag for disse oppgavene og viser at regional ulikhet påvirkes av flere strukturelle forhold samtidig.

I oppgave 1 benyttes Cappelen & Mjøset (2009) til å gi ytterligere forklaringer på hvordan regionale økonomiske forskjeller oppstår. Cappelen & Mjøset (2009) bruker ressursforvaltning

og økonomisk utvikling i Norge, og viser at naturressurser, næringsstruktur og institusjonell utforming påvirker regioners økonomiske utviklingsmuligheter. Dette innebærer at ressursrike regioner kan ha en annen vekstbane enn andre regioner, noe som igjen påvirker regional utvikling. Videre gir Bandeira Morais et al. (2021) en presis introduksjon til Gini-koeffisienten, og understreker at Gini er et av de mest brukte målene for økonomisk ulikhet, der verdien går fra 0 (full likhet) til 1 (full ulikhet). Dette gir et teoretisk grunnlag for å bruke Gini-koeffisienten som hovedindikator for regional inntektsulikhet i disse oppgavene.

I oppgave 2 brukes utdanning, infrastruktur og demografi som sentrale forklaringsvariabler for regional ulikhet. Coady & Dizioli (2017) viser at utdanningsulikhet har en positiv sammenheng med inntektsulikhet, og at økte utdanningsmuligheter kan redusere forskjeller. Dermed vil regioner med høyere utdanningsnivå ofte ha lavere ulikhet. Når det gjelder infrastruktur, finner Calderón & Servén (2004) at forbedringer i blant annet veitethet bidrar til økonomisk vekst og reduserer regional ulikhet, og at veitethet typisk har en negativ sammenheng med Gini. Innen demografi viser Dolls et al. (2019) at aldring kan øke ulikhet, ettersom en høyere andel eldre kan bidra til større inntektsforskjeller.

I oppgave 3 viser Kuznets (2019) utviklingen i forholdet mellom økonomisk vekst og inntektsfordeling, og argumenterer for at ulikhet kan følge et mønster der den først øker, deretter stabiliserer seg og til slutt faller, med vekt på at ulikhetsnivået kan endres med økonomiske strukturer og utviklingsstadier.

Samlet sett bidrar disse litteraturene, gjennom perspektiver på regionale økonomiske forskjeller, målemetoder for ulikhet og faktorer som utdanning, infrastruktur og demografi, til det teoretiske utgangspunktet for disse oppgavene. Litteraturen peker gjennomgående på at ulikhet påvirkes av flere strukturelle forhold, og at disse kan variere betydelig mellom regioner. De nevnte litteraturene danner et faglig grunnlag for å analysere forholdet mellom regional utvikling og inntektsulikhet.

## Forskningshull

Selv om eksisterende litteratur gir et viktig teoretisk grunnlag for å forstå forholdet mellom regional utvikling og inntektsulikhet, finnes det fortsatt områder som er lite utforsket.

For det første, på datanivå, bygger tidligere arbeid som Lessmann & Seidel (2017) hovedsakelig på nasjonale eller globale sammenligninger og analyserer sammenhengen mellom utvikling og ulikhet ved hjelp av tverrnasjonale datasett. Artikkelen benytter ikke Eurostat sine harmoniserte og direkte sammenlignbare NUTS2-regiondata innenfor samme statistiske rammeverk.

For det andre, fokuserer mange artikler ofte på en enkelt faktor, som utdanning, infrastruktur eller demografi, og analyserer sjeldent disse variablene samtidig innenfor samme empiriske ramme. I disse oppgavene analyseres imidlertid utdanningsnivå (høy utdanning), transportinfrastruktur (veitethet), demografisk struktur (65 år eller eldre) og økonomisk utvikling (BNP per innbygger eller endringen i BNP per innbygger) samlet i en modell i forhold til regional inntektsulikhet (Gini).

For det tredje, bruker tidligere arbeid en enkelt analysemodell, enten tverrsnittsanalyser eller tidsseriestudier, og kombinerer sjeldent flere tilnærmingar som langtidsbeskrivelser, tverrsnittsregresjoner og paneldataanalyse. Dermed blir robuste tester av de samme sammenhengene på tvers av ulike metoder ofte fraværende.

Disse oppgavene benytter Eurostat-data og gjennomfører en sammenlignende analyse av Belgia, Bulgaria, Nederland og Norge. Gjennom langtidsbeskrivelsen av BNP per innbygger i oppgave 1, tverrsnittsanalysen i oppgave 2, og de alternative funksjonelle formene samt panelmodellene i oppgave 3, analyseres sammenhengene mellom regional økonomisk utvikling (BNP per innbygger eller endringen i BNP per innbygger), utdanning (høy utdanning), transportinfrastruktur (veitetthet), demografisk struktur (65 år eller eldre) og regional inntektsulikhet (Gini) fra flere perspektiver. På denne måten gir oppgavene en helhetlig tilnærming der de samme variablene undersøkes fra ulike metodiske vinkler, basert på én konsistent datakilde.

## Data og metodikk

### Datakilder

Datasettet som brukes i oppgave 1 kommer fra Eurostat og inkluderer «nama\_10r\_3gdp» (BNP) og «demo\_r\_pjanaggr3» (befolkningsdata), begge tilgjengelig på NUTS3-nivå. Basert på BNP- og befolkningsdata for perioden 2000–2023 beregnes BNP per innbygger (BNPC), og flere NUTS3-regioner slås sammen til NUTS2 ved hjelp av befolkningsvektning. Innenfor samme sammenslåingsramme beregnes også Gini-koeffisienten på NUTS2-nivå for å måle regional inntektsulikhet. Siden Eurostat kun dekker Norge fra og med 2010, er tilgjengelige år for Norge kortere enn for de andre landene i disse oppgavene. I tillegg, selv uten å slette manglende verdier, varierer antall tilgjengelige observasjoner mellom land i de opprinnelige Eurostat-datasettene, og sammenslåing etter region fører også til tap av enkelte observasjoner. Noen NUTS2-regioner består kun av en NUTS3-enhet, noe som genererer flere NA-verdier ved beregning av Gini.

Oppgave 2 benytter 2017-data hentet fra datasett brukt i oppgave 1, og henter i tillegg nye datasett fra Eurostat: utdanning (høy utdanning, ISCED 5–8) fra «edat\_lfse\_04», demografisk struktur (andel av befolkningen som er 65 år eller eldre) fra «demo\_r\_pjanaggr3», for transportinfrastruktur hentes veilengde fra «tran\_r\_net», og arealdata fra «reg\_area3», som kombineres for å beregne veitetthet. Disse variablene utgjør 2017 tverrsnittsgrunnlaget for analysen av regional inntektsulikhet. Datasettene varierer i antall observasjoner mellom land, og enkelte regioner har manglende verdier, noe som fører til redusert utvalgsstørrelse etter datarensing.

Datasettet som brukes i oppgave 3 dekker perioden 2008–2022 og benytter de samme variablene som i oppgave 1 og oppgave 2, hentet fra de samme Eurostat-kildene. Det konstrueres analysefiler på tvers av år og modeller basert på samme datagrunnlag. Siden tilgjengelige år og datadekning varierer mellom land, får panelet en ubalansert struktur (unbalanced panel), og enkelte regioner mangler observasjoner i enkelte år. Dette reduserer antallet gyldige årsobservasjoner som kan inngå i estimeringen.

### Metodisk tilnærming

Oppgave 1 benytter en beskrivende metode (beskrivende analyse) ved å beregne og sammenligne BNP per innbygger (BNPC) og Gini-koeffisienten på NUTS2-nivå for perioden 2000–2023. Dette brukes for å observere og sammenligne regional utvikling og ulikhet i Belgia, Bulgaria, Nederland og Norge. Gjennom deskriptiv statistikk og visualisering, der blant annet figurene fordeling av Gini (NUTS2) etter land og Gini per NUTS2 over tid (fasettert etter land), synliggjøres ulikhetsfordelingen mellom regioner og utviklingen over tid på en tydelig måte.

Den beskrivende analysen er hensiktsmessig som et første steg før modellering, ettersom den gir en oversikt over langsiktige mønstre i regional økonomisk utvikling og identifiserer sentrale forskjeller mellom regioner.

Oppgave 2 bygger på tverrsnittsdata for 2017 og estimerer både en enkel lineær regresjonsmodell og en multiple lineær regresjonsmodell. I den enkle lineære regresjonsmodellen testes hovedantakelsene i OLS, inkludert linearitet, uavhengighet, homoskedastisitet, fravær av multikollinearitet og normalfordeling av residualer. Deretter inkluderes nye forklaringsvariabler i den multiple modellen, blant annet høyere utdanning, veitethet og andelen av befolkningen som er 65 år eller eldre. Tverrsnittstilnærmingen gjør det mulig å kontrollere flere faktorer på samme tidspunkt og sammenligne ulikhetsnivåer mellom regioner, noe som gir grunnlag for å analysere statistiske sammenhenger mellom forklaringsvariabler og regional inntektsulikhet.

Oppgave 3 benytter paneldata for perioden 2008–2022 og inkluderer delmengdeanalyse, alternative funksjonelle former som log-lineær modell, kvadratisk modell og kubisk modell, heteroskedastisitetstesting, samt panelestimering (region FE, year FE, two-way FE og country FE). De alternative funksjonelle formene brukes for å undersøke om forholdet mellom variabler er robust under ulike modellspecifikasjoner. Deretter estimeres faste effekter (FE) for å kontrollere for region-, år- eller lands-spesifikke egenskaper som ikke varierer over tid, slik at modellen utnytter variasjon på tvers av år for å estimere sammenhengen mellom økonomisk utvikling, høyere utdanning, veitethet og andelen av befolkningen som er 65 år eller eldre og regional inntektsulikhet (Gini). Fordelen med panel-FE-modeller er at de reduserer skjevhetsfaktorer og dermed gir mer robuste estimater.

## Empiriske funn

### Resultater fra oppgave 2

#### Enkel lineær regresjonsmodell

Tabell 1: Resultater fra enkel regresjon.

	Estimat	Standard Error	t value	Pr(> t )
(Intercept)	0.033	0.019	1.748	0.0951 .
change_GDPC_pct	0.010	0.003	2.906	0.0084 **

*Signif. codes: 0 <= '\*\*\*\*' < 0.001 < '\*\*' < 0.01 < '\*' < 0.05*

Residual standard error: 0.04598 on 21 degrees of freedom

Multiple R-squared: 0.2869, Adjusted R-squared: 0.2529

F-statistic: 8.447 on 21 and 1 DF, p-value: 0.0084

Basert på tverrsnittsanalysen i oppgave 2 estimeres først en enkel lineær regresjonsmodell for å undersøke effekten av økonomisk utvikling (endringen i BNP per innbygger) på regional inntektsulikhet (Gini). Resultatene i Tabell 1 viser en signifikant positiv sammenheng mellom økonomisk utvikling og ulikhet. Koeffisienten for økonomisk utvikling er 0.0097 og signifikant på 1 %-nivå ( $p = 0.0084$ ), noe som innebærer at regioner med høyere økonomisk vekst ofte har høyere Gini-koeffisient. Den justerte  $R^2$ -verdien er 0.2529, som viser at økonomisk vekst alene har begrenset forklaringskraft for regionale ulikheter.

## Multipel lineær regresjonsmodell

Tabell 2: Resultater fra multipel regresjon.

	Estimat	Standard Error	t value	Pr(> t )
(Intercept)	-0.014	0.117	-0.121	0.9052
change_GDPC_pct	0.011	0.006	2.038	0.0574
utdanning_hoy	0.004	0.001	2.915	0.0097 **
vei_tetthet	-0.001	0.000	-2.232	0.0393 *
andelen_over65	-0.003	0.004	-0.663	0.5163

Signif. codes: 0 <= \*\*\*\* < 0.001 < \*\*\* < 0.01 < \*\* < 0.05

Residual standard error: 0.03782 on 17 degrees of freedom

Multiple R-squared: 0.5861, Adjusted R-squared: 0.4888

F-statistic: 6.019 on 17 and 4 DF, p-value: 0.0033

I den multiple lineære regresjonsmodellen inkluderes flere forklaringsvariabler, blant annet transportinfrastruktur (veitetthet), utdanning (høy utdanning) og demografisk struktur (andel av befolkningen som er 65 år eller eldre). Modellen gjengitt i Tabell 2 viser en tydelig forbedring i forklaringskraft (justert  $R^2 = 0.4888$ ), noe som indikerer at strukturelle forskjeller mellom regioner spiller en viktig rolle for regional inntektsulikhet. Resultatene viser at høy utdanning har en signifikant positiv effekt (koeffisient 0.00433,  $p = 0.00966$ ), som innebærer at regioner med høyere utdanningsnivå også har større inntektsforskjeller. Veitetthet har en signifikant negativ sammenheng med ulikhet (koeffisient -0.0009,  $p = 0.03934$ ), noe som viser at regioner med bedre transportinfrastruktur som regel har lavere ulikhet. Variabelen økonomisk utvikling beholder en svak positiv sammenheng i den multiple modellen (koeffisient 0.01126,  $p = 0.05738$ ), men uten sterkt statistisk signifikans. Andelen eldre over 65 år viser ingen signifikant effekt ( $p = 0.51634$ ).

Samlet sett viser tverrsnittsanalysen at regional inntektsulikhet ikke kan forklares av økonomisk utvikling alene. Selv om den enkle lineære modellen indikerer en positiv sammenheng mellom økonomisk vekst og ulikhet, forklarer modellen kun om lag en fjerdedel av variasjonen mellom regioner. Den utvidede multiple modellen har derimot betydelig høyere forklaringskraft og viser at utdanningsnivå og transportinfrastruktur har de sterkeste og mest konsistente sammenhengene med ulikhet, noe som er klart sterkere enn økonomisk vekst. Sammenligningen av de to modellene viser at regional ulikhet er et resultat av flere strukturelle forhold, der utdanning og transportinfrastruktur spiller en langt viktigere rolle enn økonomisk utvikling alene.

## Resultater fra oppgave 3

### Alternative funksjonelle former

Tabell 3: logaritmiske, kvadratiske, kubiske

	Log-lineær modell	Kvadratisk modell	Kubisk modell
(Intercept)	-0.12151 (-0.93697)	-0.14363 (-1.02523)	-0.07947 (-0.46799)
log(change_GDPC_pct + 1)	0.08266* (2.36393)		
vei_tetthet	-0.00087* (-2.24811)	-0.00086* (-2.20387)	-0.00078+ (-1.87064)
andelen_over65	-0.00236 (-0.57973)	-0.00046 (-0.10355)	-0.00003 (-0.00574)
Høy_utdanning	0.00467** (3.17124)	0.00475** (3.26921)	0.00450** (2.95712)
change_GDPC_pct		0.03697* (2.12585)	-0.00154 (-0.02643)
I(change_GDPC_pct^2)		-0.00200 (-1.55222)	0.00403 (0.45855)
I(change_GDPC_pct^3)			-0.00028 (-0.69392)
Num.Obs.	22	22	22
R2	0.612	0.640	0.651
R2 Adj.	0.521	0.528	0.512
AIC	-76.8	-76.4	-75.1
BIC	-70.2	-68.8	-66.4
Log.Lik.	44.387	45.209	45.557
F	6.715	5.696	4.673
RMSE	0.03	0.03	0.03

+ p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

Viser t-verdier i parentes under estimatene.

I Tabell 3 presenteres tre alternative funksjonelle former i oppgave 3: log lineær modell, kvadratisk modell og kubisk modell. Modellene estimeres med de samme kontrollvariablene som i oppgave 2, og resultatene er oppsummert i tabell Tabell 3.

For det første viser de tre modellene R2 verdier på henholdsvis 0.612 for log modellen, 0.640 for den kvadratiske modellen og 0.651 for den kubiske modellen. Den justerte R2 ligger i området 0.512 til 0.528, med små forskjeller mellom modellene. Når det gjelder AIC, er log modellen lavest med verdien -76.8, tett fulgt av den kvadratiske modellen med -76.4, mens den kubiske modellen er noe høyere med -75.1. Samlet sett fremstår log modellen som den beste etter modellvalgsindikatorene, mens den kvadratiske modellen har noe høyere forklaringskraft.

For det andre viser den log lineære modellen en koeffisient på 0.08266 for log(change\_GDPC\_pct + 1), som er statistisk signifikant og indikerer en positiv sammenheng mellom økonomisk utvikling og regional ulikhet. I den kvadratiske modellen er koeffisienten for change\_GDPC\_pct 0.03697 og fortsatt signifikant, men kvadratet av variabelen har ingen signifikant effekt. I den kubiske modellen er ingen av de tre vekstrelaterte parameterne, change\_GDPC\_pct, I(change\_GDPC\_pct^2) eller I(change\_GDPC\_pct^3), signifikante. Dette

viser at inkludering av tredjegradsleddet ikke gir ekstra forklaringskraft, og modellen preserer svakere enn både log og kvadratisk modell.

For det tredje er variabelen *høy\_utdanning* signifikant og positiv i alle tre modeller, med koeffisienter rundt 0.00475 til 0.0045. Dette viser at utdanningsnivået er den mest stabile forklaringsvariablene i modellene. Variablen *vei\_tetthet* har negativ effekt i alle tre modellene og er signifikant i både log og kvadratiske modeller, med koeffisienter på henholdsvis -0.00087 og -0.00086. I den kubiske modellen er effekten svakere og kun svakt signifikant, med koeffisient -0.00078. Variablen *andelen\_over65* er ikke signifikant i noen av modellspefikasjonene.

Samlet sett viser tabell Tabell 3 at log lineær, kvadratisk og kubisk modell alle fanger opp sammenhengen mellom økonomisk utvikling og inntektsulikhet, men med ulike resultater. Log lineær modell og kvadratisk modell viser tydeligere statistiske sammenhenger for vekstvariablene, mens den kubiske modellen, til tross for høy  $R^2$ , ikke har signifikante vekstparametere og tilfører lite informasjon. Basert på sammenligning av  $R^2$  og AIC vurderes log lineær og kvadratisk modell som de mest representative alternative funksjonelle formene.

### Panelmodeller (Fire FE-spesifikasjoner)

Tabell 4: Sammenligning av panelmodeller

	Region FE	Year FE	Two-way FE	Country FE
Høy_utdanning	-0.00036 (0.00044)	0.00011 (0.00061)	-0.00035 (0.00053)	0.00041 (0.00084)
andelen_over65	0.00048 (0.00088)	0.00315+ (0.00185)	0.00096 (0.00125)	-0.00055 (0.00155)
Num.Obs.	204	204	204	35
R2	0.004	0.022	0.006	0.009
R2 Adj.	-0.130	-0.028	-0.180	-0.162
AIC	-1242.3	-627.2	-1259.9	-247.4
BIC	-1232.3	-617.3	-1249.9	-242.8
RMSE	0.01	0.05	0.01	0.01

+ p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

For å analysere de dynamiske endringene innenfor regioner over tid estimerer oppgave 3 fire typer faste effektsmodeller: region FE, year FE, two way FE og country FE. Variablen *veitethet* er ikke inkludert i panelmodellene, siden faste effekter kun kan estimere variable som varierer over tid innen samme region. Variable som er tidsinvariante eller har svært liten variasjon kan ikke få identifisert koeffisienter. Derfor inneholder panelanalysen kun variablene *høy\_utdanning* og *andelen\_over65*, som begge varierer over tid.

For det første viser tabell Tabell 4 at alle panelmodellene har svært lave R2 verdier. Region FE har  $R^2$  på 0.004, year FE 0.022, two way FE 0.006 og country FE 0.009. Den justerte  $R^2$  er negativ for alle modellene i området fra -0.130 til -0.028. Dette viser at uansett hvilken type faste effekter som benyttes, har modellene svært begrenset evne til å forklare variasjonen i Gini over tid.

For det andre viser koeffisientenes signifikans at nesten ingen variabler er signifikante i noen av de fire modellene. *Høy\_utdanning* ligger svært nær null i alle spesifikasjoner, med koeffisienter mellom -0.00036 og 0.00041, og ingen av dem er signifikante. *Andelen\_over65* viser svak signifikans i year FE modellen med koeffisient 0.00315, men er ikke signifikant i de øvrige modellene. Dette betyr at både utdanningsnivå og andelen eldre over 65 år har svært liten variasjon over tid, og selv etter kontroll for faste effekter bidrar disse variablene i liten grad til å forklare årlige endringer i Gini.

For det tredje viser modellsammenligningen at year FE modellen har den høyeste  $R^2$  med 0.022, men forklaringskraften er fortsatt svært lav. Two way FE kontrollerer både region og år, men har likevel en  $R^2$  på kun 0.006, som er lavere enn year FE. Country FE modellen bruker faste effekter på landsnivå, men inkluderer bare 35 observasjoner, og ingen av koeffisientene er signifikante.

Samlet sett viser tabell Tabell 4 at alle fire panelmodellene har svært svak forklaringskraft, og at nøkkelvariablene ikke viser signifikante effekter i tidsdimensjonen. Koeffisientene ligger nær null, og  $R^2$  verdiene er gjennomgående lave. Dette viser at faste effektsmodeller ikke klarer å identifisere tydelige dynamiske endringer i regional inntektsulikhet over tid i dette datasettet.

## Diskusjon

### Viktige innsikter

Gjennom en komplementær analyse av tverrsnittsmodeller, alternative funksjonelle former og panelmodeller fremkommer det at regional inntektsulikhet har en kompleks og flerlagd struktur, og at ulikhet ikke bestemmes av en enkelt faktor, men av samspillet mellom flere regionale kjennetegn.

For det første viser den enkle lineære regresjonsmodellen en positiv sammenheng mellom økonomisk utvikling og regional ulikhet, men modellens forklaringskraft er lav med en justert  $R^2$  på om lag 25 prosent. Dette viser at økonomisk vekst alene ikke er tilstrekkelig for å forklare ulikhetsnivået mellom regioner.

For det andre viser den multiple lineære regresjonsmodellen en betydelig økning i forklaringskraft, til rundt 49 prosent, når sentrale regionale faktorer som utdanning (høy utdanning), transportinfrastruktur (veitetthet) og demografisk struktur (andel av befolkningen som er 65 år eller eldre) inkluderes. Høy utdanning har en signifikant positiv effekt, noe som viser at utdanning er den viktigste forklaringsvariabelen. Veitetthet har en signifikant negativ effekt, noe som innebærer at regioner med bedre transportinfrastruktur har lavere inntektsulikhet. Økonomisk utvikling beholder en positiv koeffisient, men uten statistisk signifikans, noe som viser at dens forklaringsbidrag er begrenset.

For det tredje viser de alternative funksjonelle formene at sammenhengen mellom økonomisk utvikling og regional ulikhet varierer noe i signifikans på tvers av modellspesifikasjoner. Selv om log modellen og den kvadratiske modellen gir bedre modelltilpasning, er hovedresultatene stabilt: høy utdanning og veitetthet er signifikante i alle modeller, mens økonomisk utvikling viser svakere og mindre stabile effekter. Dette viser at utdanning og infrastruktur gir mer robuste forklaringer på ulikhet enn økonomisk vekst.

Til slutt viser panelmodellene at alle faste effektsmodeller har svært lav forklaringskraft. Verken country FE, year FE, region FE eller two way FE viser signifikante effekter for nøkkelvariablene over tid. Dette gjenspeiler at regional ulikhet, utdanningsnivå og befolningsstruktur i perioden 2008 til 2022 endrer seg svært lite innen regioner, noe som gjør det vanskelig for faste effekter å identifisere tydelige tidsvariasjoner.

Samlet sett viser disse modellene at forskjeller i regional inntektsulikhet hovedsakelig kommer fra langsiktige forskjeller mellom regioner, særlig knyttet til utdanningsnivå og transportinfrastruktur, og ikke fra endringer over tid. Høy utdanning og veitetthet er signifikante i både tverrsnittsmodeller og de alternative funksjonelle formene, mens variablene ikke viser signifikante effekter i panelmodellene fordi de endrer seg svært lite over tid. Dette viser at regional inntektsulikhet i stor grad gjenspeiler varige forskjeller mellom regioner, og ikke variasjoner fra år til år.

## **Politiske implikasjoner**

Resultatene fra oppgavene viser at utdanningsnivå og transportinfrastruktur er de mest konsekvente faktorene knyttet til regional inntektsulikhet, mens økonomisk vekst viser svakere og mindre stabile sammenhenger. Dette gir et grunnlag for å vurdere hvilke typer tiltak som kan være mest målrettet når en ønsker å redusere regionale forskjeller.

For det første er utdanningsnivået konsekvent signifikant i både tverrsnittsmodellen og de alternative funksjonelle formene, og fremstår som den viktigste forklaringsvariabelen. Dette innebærer at økt tilgang til utdanning er ett av de mest effektive langsiktige virkemidlene for å redusere regional ulikhet. I regioner med svake utdanningsressurser bør politikken prioritere bedre tilgang til høyere utdanning og kompetanseheving, for å styrke regionale utviklingsmuligheter.

For det andre viser transportinfrastruktur, målt ved veitetthet, en tydelig negativ sammenheng med ulikhet på tvers av flere modeller. Dette tyder på at bedre regional tilknytning kan fremme balansert utvikling. For regioner med svak infrastruktur kan investeringer i transportforbindelser være et viktig virkemiddel for å redusere regionale forskjeller.

For det tredje er den økonomiske veksten ustabil i sin signifikans, og forklaringskraften er betydelig svakere enn for utdanning og infrastruktur. Dette betyr at økonomisk vekst alene ikke er et effektivt virkemiddel for å redusere regionale forskjeller. Regionalpolitikken bør derfor rette større oppmerksomhet mot strukturelle faktorer som utdanning og infrastruktur, i stedet for kortsiktige vekststimulerende tiltak.

Til slutt viser panelanalysen at de interne endringene innen regioner i perioden 2008 til 2022 er svært begrensete, og at regionale forskjeller har en tydelig langsiktig karakter. Dette innebærer at kortsiktige tiltak vanskelig kan endre ulikhetsmønstre raskt, og at politikken derfor må bygge på langsiktige og vedvarende strukturelle tiltak.

Med dette som bakgrunn kan politikk for å redusere regionale forskjeller med fordel prioritere tiltak som styrker utdanningsmuligheter og forbedrer transportinfrastruktur. Slike tiltak har større sannsynlighet for å gi varige forbedringer enn kortsiktige økonomiske vekststimuli.

# Begrensninger og fremtidig forskning

## Forskningsbegrensninger

Oppgavene har flere begrensninger knyttet til data og metode. For det første varierer Eurostats dekning mellom land, og sammenslåingen av NUTS3 til NUTS2 medfører enkelte manglende verdier, noe som reduserer antall observasjoner. Dette begrenser regresjonsmodellene og fører til en ubalansert panelstruktur, som igjen svekker muligheten til å identifisere tidsmessige endringer.

For det andre endrer de sentrale forklaringsvariablene seg svært lite over tid innen regionene. Dette gjelder blant annet utdanningsnivå, veitethet og demografisk struktur. Dette gjør det vanskelig for faste effektsmodeller å identifisere deres tidsdynamiske effekter, noe som fører til generelt svak signifikans i panelanalysene.

For det tredje inkluderer oppgavene kun utdanningsnivå, veitethet og demografisk struktur, ikke andre faktorer som kan påvirke regional ulikhet. På grunn av det begrensede antallet forklaringsvariabler kan modellene bare delvis belyse de faktorene som driver regional ulikhet.

Til slutt bygger oppgavene sine konklusjoner på statistiske sammenhenger observert i modellene, og forklaringskraften er begrenset av datadekning og modellspesifikasjoner. Mer detaljerte mekanismer mellom variablene krever videre forskning med bredere datagrunnlag og utvidede metodiske tilnærminger.

# Konklusjon

## Sammendrag

Gjennom å kombinere den langsiktige deskriptive analysen i oppgave 1, tverrsnittsregresjonsmodellen i oppgave 2 og de alternative funksjonelle formene samt panelmodellene med faste effekter i oppgave 3, undersøker de oppgavene systematisk regionale forskjeller i økonomisk utvikling og sentrale faktorer som påvirker regional inntektsulikhet i Belgia, Nederland, Bulgaria og Norge.

Tverrsnittsanlysen viser at høy utdanning og veitethet har stabile og signifikante effekter på tvers av ulike modellspecifikasjoner, og er de viktigste forklaringsfaktorene for regional ulikhet. Høy utdanning har en signifikant positiv sammenheng med inntektsulikhet, noe som innebærer at regioner med en større andel høyt utdannede ofte har høyere ulikhet. Veitethet har derimot en negativ sammenheng med ulikhet, og regioner med bedre transporttilgang har som regel lavere inntektsforskjeller. Sammenlignet med dette viser endringen i BNP per innbygger ustabile signifikansnivåer i flere modeller, noe som tyder på at økonomisk vekst ikke er den viktigste forklaringsfaktoren bak regional ulikhet.

Resultatene fra de alternative funksjonelle formene bekrefter dette bildet. Selv om log lineære og kvadratiske modeller gir en viss forbedring i modelltilpasning, endres verken koeffisientenes retning eller deres signifikans på en vesentlig måte. Effekten av høy utdanning og veitethet forblir robust, mens økonomisk vekst fortsatt har begrenset forklaringskraft.

Panelmodellene viser at de sentrale variablene har svært begrenset tidsvariasjon innen regionene i perioden 2008 til 2022. Høy utdanning og den demografiske strukturen (andel 65 år eller eldre) fremstår uten signifikante effekter i faste effektsmodeller. Dette indikerer at

regional inntektsulikhet i større grad gjenspeiler vedvarende forskjeller mellom regioner enn kortsiktige årlige endringer.

Samlet sett viser modellresultatene at tiltak for å redusere regionale inntektsforskjeller bør legge større vekt på langsiktige innsatsområder som å øke utdanningsnivå og forbedre transportinfrastruktur, i stedet for å støtte seg på kortsiktige tiltak for å stimulere økonomisk vekst.

Betydningen av disse funnene er at de bidrar til å identifisere hvilke faktorer som har sterkest sammenheng med regional ulikhet, noe som kan hjelpe beslutningstakere med å målrette innsatsen mer effektivt.

### **Sluttrefleksjon**

Basert på de ulike datastrukturene og metodiske tilnærmingene som oppgave 1, 2 og 3 benytter, har forholdet mellom regional utvikling og inntektsulikhet kunnet undersøkes fra flere perspektiver. Ved å kombinere deskriptive analyser, tverrsnittsmodeller og panelestimeringer har oppgavene vist hvordan valg av metode og datagrunnlag påvirker de empiriske resultatene, og understreket betydningen av å bruke flere metodiske tilnærninger i regional forskning.

Utfordringer knyttet til datadekning på tvers av land, manglende observasjoner og begrenset tidsvariasjon i sentrale variabler gjenspeiler også praktiske begrensninger som ofte oppstår i analyser av regionale data. Til tross for disse utfordringene viser resultatene på tvers av modellene en viss konsistens, noe som gjør det mulig å identifisere utdanningsnivå og transportinfrastruktur som de mest sentrale faktorene knyttet til regionale forskjeller.

Analysene gir et mer konkret bilde av hvordan et lite sett av variabler kan være knyttet til regional inntektsulikhet. Selv om analysene ikke kan si noe om årsakssammenhenger, gir arbeidet nyttige erfaringer med bruk av ulike metoder, noe som kan være relevant for videre empiriske studier på området.

### **Vedlegg-Bruken av AI-verktøy**

Jeg benyttet ChatGPT (modell GPT-5.1) i arbeidet med MSB104 oppgave 4, hovedsakelig som støtte til oversettelse mellom kinesisk og norsk. Basert på mine egne kinesiske utkast brukte jeg ChatGPT til å oversette innholdet til norsk og til å få forslag til mer naturlige eller mer akademiske formuleringer. Jeg gjennomgikk alle forslagene og teksten som ble generert av ChatGPT, redigerte og omorganiserte dem for å sikre at terminologien, strukturen og det analytiske rammeverket var i samsvar med det opprinnelige utkastet. Gjennom hele prosessen bidro ChatGPT til å forbedre sammenheng og språklig klarhet i tekstene. Alle formuleringer ble kontrollert av meg setning for setning, og jeg sørget for at de var i tråd med min egen analyse og forståelse.

### **Referanser**

Bandeira Morais, M., Swart, J., & Jordaan, J. A. (2021). Economic Complexity and Inequality: Does Regional Productive Structure Affect Income Inequality in Brazilian States? *Sustainability*, 13(2). <https://doi.org/10.3390/su13021006>

Calderón, C., & Servén, L. (2004). *The Effects of Infrastructure Development on Growth and Income Distribution* (Policy Research Working Paper Nr. 3400). World Bank. <https://hdl.handle.net/10986/14136>

Cappelen, Ø., & Mjøset, L. (2009). Can Norway Be a Role Model for Natural Resource Abundant Countries? *Discussion Papers in Economics*, 09, 1–26. [https://www.researchgate.net/publication/228983025\\_Can\\_Norway\\_Be\\_a\\_Role\\_Model\\_for\\_Natural\\_Resource\\_Abundant\\_Countries](https://www.researchgate.net/publication/228983025_Can_Norway_Be_a_Role_Model_for_Natural_Resource_Abundant_Countries)

Coady, D., & Dizioli, A. (2017). Income Inequality and Education Revisited: Persistence, Endogeneity, and Heterogeneity. *IMF Working Papers*, 17(126), 1. <https://doi.org/10.5089/9781475595741.001>

Dolls, M., Doorley, K., Paulus, A., Schneider, H., & Sommer, E. (2019). Demographic change and the European income distribution. *The Journal of Economic Inequality*, 17(3), 337–357.

Kuznets, S. (2019). Economic growth and income inequality. I *The gap between rich and poor* (s. 25–37). Routledge.

Lessmann, C., & Seidel, A. (2017). Regional inequality, convergence, and its determinants – A view from outer space. *European Economic Review*, 92, 110–132. <https://doi.org/https://doi.org/10.1016/j.eurocorev.2016.11.009>