

Sebastian Babusiaux, Mathias Fossen, Kornelius Fuglem og Sigrid Osland

Empirisk analyse av forholdet ulikhet og økonomisk vekst

En analyse av inntektsulikhetens påvirkning på økonomisk vekst

Bacheloroppgave i Samfunnsøkonomi

Veileder: Irmelin Slettemoen Helgesen

Mai 2023

Sebastian Babusiaux, Mathias Fossen, Kornelius
Fuglem og Sigrid Osland

Empirisk analyse av forholdet ulikhet og økonomisk vekst

En analyse av inntektsulikhetens påvirkning på
økonomisk vekst

Bacheloroppgave i Samfunnsøkonomi
Veileder: Irmelin Slettemoen Helgesen
Mai 2023

Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
Institutt for samfunnsøkonomi



Kunnskap for en bedre verden

Forord

Det følgende arbeidet er skrevet som en konkluderende oppgave for vår bachelorgrad ved avdeling for økonomi ved Norges Teknisk-Naturvitenskapelige Universitet.

Vi ønsker å uttrykke ekstra takknemlighet overfor veilederen vår, Irmelin Slettemoen Helgemoen, for den støtten og hjelpen vi har fått gjennom prosessen å skrive oppgaven.

Sammendrag

Inntektsulikhet er en faktor som kan påvirke direkte økonomiske størrelser, som for eksempel økonomisk vekst. Inntektsulikhets påvirkning på økonomisk vekst er uklar. Det finnes teorier som forfekter en positiv påvirkning, og det finnes teorier som forfekter en negativ påvirkning. Denne oppgaven analyserer forholdet mellom inntektsulikhet og økonomisk vekst, og skal vurdere om det foreligger ulike sammenhenger for lavinntektsland og høyinntektsland.

Datasettet i oppgaven er hentet fra verdensbanken, og analysen anvender minste kvadraters metode (OLS) som økonometrisk metode. Vi gjennomfører to ulike regresjonsanalyser, en for lavinntektsland og en for høyinntektsland. I analysen benyttes det tidsseriedata fra 1960-2021 for de to utvalgene. Modellen legger til kontrollvariabler en etter en for å undersøke effekten inntektsulikhet har på økonomisk vekst.

Resultatene fra analysen er tvetydig. For både lavinntektsland og høyinntektsland er funnene statistisk signifikante, men inntektsulikhet synes å påvirke den økonomiske veksten ulikt. For lavinntektsland viser resultatene en konkav sammenheng, mens for høyinntektsland er sammenhengen konveks. Dette vil si at for lavinntektsland har inntektsulikhet en avtakende positiv effekt på økonomisk vekst, fram til et toppunkt. Etter dette toppunktet vil inntektsulikheten påvirke den økonomiske veksten negativt. Derimot tilsier resultatene for høyinntektsland at inntektsulikhet har en avtakende negativ effekt på økonomisk vekst, frem til et bunnpunkt. Etter dette punktet vil inntektsulikhet påvirke økonomisk vekst positivt.

Nøkkelord: ulikhet, økonomisk vekst, tids-serie data, hypotese testing

Innholdsfortegnelse

1. Innledning	6
2. Teoretisk rammeverk og litteratur	7
2.1 Klassisk tilnærming	7
2.2 Moderne tilnærminger	8
2.2.1 Kuznets og Piketty	8
<i>Figur 1: Kuznets-kurven</i>	9
<i>Figur 2: Invers Kuznets-kurve</i>	9
2.2.2 Effekten gjennom den politiske og sosiopolitiske kanalen	10
3. Data	11
3.1.1 Avhengig variabel	12
3.1.2 Forklaringsvariabel	13
<i>Figur 3: Lorenz-kurven</i>	13
3.1.3 Valg av kontrollvariabler	14
3.2 Deskriptiv statistikk	16
<i>Tabell 1: Deskriptiv statistikk for hele datasettet</i>	17
<i>Tabell 2: Deskriptiv statistikk for lavinntektsland og høynntektsland</i>	18
4. Metode	19
4.1 Forutsetninger for minste kvadraters metode	21
4.1.1 MLR.1 – Linearitet i parametere	21

4.1.2 MLR.2 – Tilfeldig utvalg	21
4.1.3 MLR.3 – Ingen perfekt kollinearitet.....	21
4.1.4 MLR.4 – Null-betinget gjennomsnitt	22
4.1.5 MLR.5 – Homoskedastisitet.....	22
4.1.6 MLR.6 – Normalitetsforutsetningen	22
4.2 Passform.....	23
4.3 Hypotesetesting	24
5. Regresjonsanalyse	25
5.1 Modell 1: for lavinntektsland	26
<i>Tabell 3: Regresjonen for lavinntektsland</i>	<i>27</i>
5.2 Modell 2: for høyinntektsland	31
<i>Tabell 4: Regresjonen for høyinntektsland</i>	<i>32</i>
5.3 Modell 3: Interaksjonsbegrep.....	35
<i>Tabell 5: Modell med interaksjonsbegrep.....</i>	<i>36</i>
6. Gjennomgang av forutsetningene for minste kvadraters metode.....	38
6.1.1 MLR.1 - Linearitet	38
<i>Tabell 6 og 7: Linearitet i variablene</i>	<i>39</i>
6.1.2 MLR.2 - Tilfeldig utvalg.....	40
6.1.3 MLR.3 - Ingen perfekt kollinearitet	40
<i>Tabell 8: korrelasjons matrise lavinntektsland.....</i>	<i>41</i>
<i>Tabell 9: VIF for lavinntektsland</i>	<i>42</i>

<i>Tabell 10: Korrelasjonsmatrise høyinntektsland</i>	42
<i>Tabell 11: VIF for høyinntektsland</i>	43
6.1.4 MLR.4 – Null-betinget gjennomsnitt	43
6.1.5 MLR.5 - Homoskedasitet	43
6.1.6 MLR.6 – Normalitetsantakelsen	44
<i>Figur 7: Normal fordeling for lav- og høyinntekts land</i>	45
7. Resultater og diskusjon	46
8. Robusthet og kritikk	47
9. Konklusjon	50
10. Litteraturliste	51

1. Innledning

De siste tre-fire tiårene har de relative forskjellene mellom rike og fattige land blitt mindre, mye på grunn av at de fremvoksende landene har hatt høyere vekst enn de rike. Samtidig øker de økonomiske ulikhetene internt i flere land i verden (NHO, 2018). Det finnes åpenbare etiske problemstillinger knyttet til økte ulikheter i et land, som for eksempel at ingen fortjener å leve i absolutt fattigdom samtidig som andre lever i overflod. Utover de etiske problemstillingene kan fordelingen av ressurser ha konsekvenser for økonomisk vekst. Den økonomiske effekten inntektsulikhet kan ha på økonomisk vekst har fått vår interesse. Årsakssammenheng mellom økonomisk vekst og inntektsulikhet er uklar, noe som kan gi problemer med simultan endogenitet i vår analyse. Likevel vil analysen prøve å skille ut inntektsulikhets påvirkning på økonomisk vekst.

Graden av ulikhet i et land kan vurderes gjennom inntektsulikheten. Det finnes flere ulike mål på inntektsulikhet. I denne analysen bruker vi de ulike landenes Gini-koeffisienter som måleenhet. Gini-koeffisienten er et praktisk og kortfattet sammendragsmål for den relative graden av inntektsulikhet i et land. Den er et aggregert ulikhetsmål, og kan variere fra 0, som utgjør perfekt likhet, til 1, som tilsier perfekt ulikhet (Todaro og Smith, 2020, s. 226).

Økonomisk vekst beskriver en økning i kvantitet og kvalitet av økonomiske varer og tjenester som et samfunn produserer og konsumerer (Roser, 2013). Det finnes flere ulike mål på økonomisk vekst. Vi har valgt å bruke BNP per innbygger som vårt mål på vekst. BNP per innbygger er en sentral makroøkonomisk hovedstørrelse som sier noe om utviklingen og verdiskapningen til et land. Fordelingen i et land påvirker den økonomiske veksten i landet gjennom blant annet landets samlede sparing, tilgangen til kreditt, samt tilliten i samfunnet.

Fagfeltet er delt, og det er flere teorier om nettopp denne relasjonen. Enkelte teorier argumenterer for at det er en positiv sammenheng mellom økt ulikhet og økt økonomisk vekst (Kaldor, 1957). På den andre siden, har vi studier som påstår det motsatte, altså at økt ulikhet fører til lavere økonomisk vekst (Persson og Tabellini (1994), Alesina og Perotti (1994)). Vi skal se nærmere på de ulike teoriene i videre diskusjon.

Denne bacheloroppgaven vil undersøke om lands forskjeller i inntektsnivå per innbygger påvirker inntektsulikhets effekt på økonomisk vekst forskjellig. Dette gjøres ved å se på følgende problemstilling:

“Har inntektsulikhet en negativ eller positiv effekt på økonomisk vekst, og vil denne sammenhengen være forskjellig for høyinntekts- og lavinntektsland?”

Oppgaven struktureres slik: Innledningsvis presenteres de ulike teoriene som allerede finnes på fagfeltet. Videre vil det redegjøres for oppgavens datagrunnlag. Her introduserer vi hvilke variabler som brukes i analysen, og avslutter med å illustrere den deskriptive statistikken til dataen. Del fire viser til hvilke økonometriske verktøy som skal brukes i videre analyse. I del fem presenteres selve regresjonsanalysen. På bakgrunn av oppgavens problemstilling vil det gjennomføres to separate regresjoner; en for lavinntektsland og en for høyinntektsland. Det vil også testes om sammenhengene er kvadratiske. Videre inkluderes det også et interaksjonsbegrep. Del seks diskuterer hvorvidt analysen oppfyller forutsetningene for minste kvadraters metode, som er presentert i del to. Avslutningsvis diskuterer analysens resultater i lys av oppgavens teorigrunnlag. Oppgavens begrensninger vil klargjøres, før konklusjonen gjør noen sluttbemerkninger.

2. Teoretisk rammeverk og litteratur

Sammenhengen mellom ulikhet og økonomisk vekst har blitt studert lenge. Det er gjennomført både teoretisk og empirisk forskning som prøver å komme fram til hvordan inntektsulikhet påvirker økonomisk vekst. Som allerede nevnt, finnes det studier med motstridende konklusjoner på dette feltet. Det vil i denne delen redegjøres for de ulike teoriene, som videre danner det teoretiske grunnlaget for oppgavens analysedel. Før vi avslutningsvis presenterer metoden som blir brukt i analysen.

2.1 Klassisk tilnærming

Den klassiske tilnærmingen viser til en positiv sammenheng mellom ulikhet og økonomisk vekst ((Keynes og Cox, 2019, s. 54), (Kaldor, 1957, s. 603)). Teorien begrunnes med at dersom inntektsulikhetene øker, og de rike blir relativt rikere, vil innbyggerne spare mer. Dette skyldes at den marginale sparetilbøyeligheten er høyere blant innbyggere som allerede har dekket sine mest grunnleggende behov. Denne tilnærmingen var spesielt viktig for den økonomiske utviklingen i den post-industrielle perioden. På den tiden var en vanlig tankegang at økt ulikhet var positivt for økonomisk vekst. Det var dermed positivt for veksten i samfunnet å kanalisere ressursøkningen mot de allerede velstående. Keynes og Kaldor

konkluderte med at økt ulikhet fører til økt samlet sparing, kapitalakkumulering, og dermed økonomisk vekst.

2.2 Moderne tilnærminger

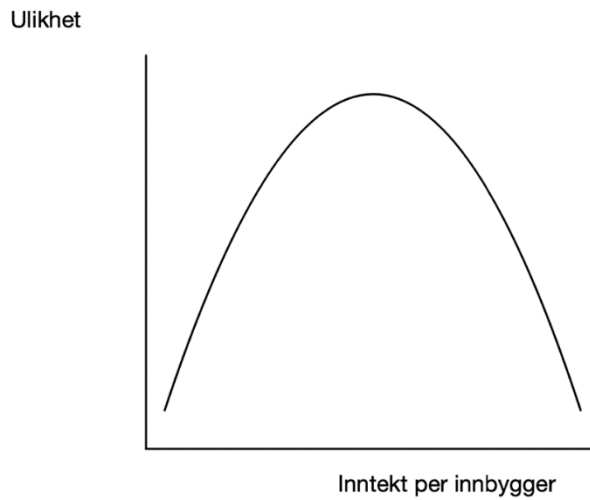
Nyere tilnærminger åpner opp for flere innfallsvinkler på forholdet mellom økonomisk vekst og inntektsulikhet. Kuznets og Piketty foreslår to ulike kvadratiske sammenhenger. Alesina og Rodrik, samt Persson og Tabellini ser på hvordan den sosio-politiske kanalen har en negativ påvirkning på forholdet. Til slutt ser Banerjee, Duflo og Galor på effekten av ulikhet gjennom et imperfekt kredittmarked.

2.2.1 Kuznets og Piketty

Som en konsekvens av moderne vekstteori ble fokuset på årsakssammenhengen mellom ulikhet og vekst endret. I 1955 argumenterte Kuznets for et motsatt årsaksforhold, hvor det er økonomisk vekst som påvirker ulikhet. Til denne teorien presenterte han historisk statistikk som viste at ulikheten i samfunnet kan forventes å følge en invers U-form, som i senere tid blir omtalt som Kuznets-kurve. Kuznets argumenterer for at det er en positiv sammenheng mellom økonomisk vekst og ulikhet på kort sikt, men at den på lang sikt derimot er negativ (Kuznets, 1955, s. 24). Han begrunnet funnet med at industrialiseringen av et land vil gi høy avkastning på kapital. Med andre ord betyr dette at avkastningen i samfunnet tilfaller kapitaleierne. En større andel av befolkningen, arbeiderne, vil oppleve at samfunnet har en økende etterspørsel etter arbeidskraft, men her vil lønningene være betydelig lavere sammenlignet med kapitaleiernes. Ergo vil ulikhetene i samfunnet øke (Roine, 2014, s. 21).

Etter hvert som industrien utvikler seg og samfunnet blir mer urbanisert, vil en stadig større andel av arbeidskraften få økt kunnskap. Dette fører til høyere produktivitet, som vil være positivt for industrien. Lønningene til arbeiderne vil i neste omgang øke, og de vil også kunne dra nytte av utviklingen. Funnene til Kuznets konkluderer med at økonomisk vekst vil føre til at inntektsforskjellene reduseres på sikt (Roine, 2014, s. 21).

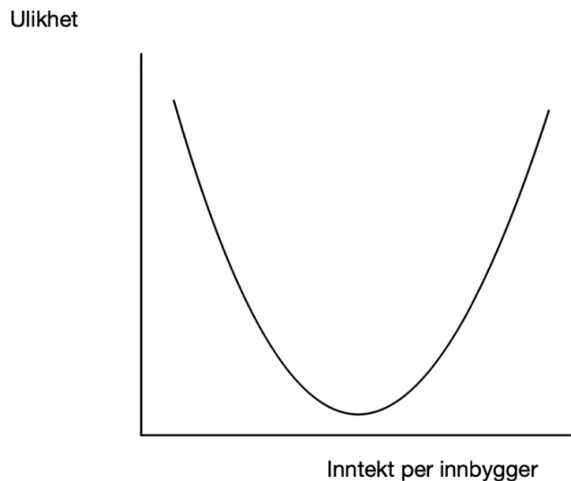
Figur 1: Kuznets-kurven



Kuznets-kurven viser hvordan økonomisk vekst kan påvirke inntektsulikhet. Usikkerhet rundt retningen i årsakssammenhengen mellom ulikhet og økonomisk vekst gir opphav til empiriske utfordringer i vår analyse, som simultan endogenitet. Dette kommer vi tilbake i diskusjonen.

Thomas Piketty utga i 2013 *Kapitalen i det 21. århundre* hvor han legger fram inntekts- og formuesstrømmer tilbake til 1700-tallet. Han sier at det ikke er noen automatikk i at inntektsforskjeller minsker i takt med økonomisk vekst (Roine, 2014, s. 21). I kontrast til Kuznets formulerer han en U-formet sammenheng mellom inntektsulikhet og økonomisk vekst.

Figur 2: Invers Kuznets-kurve



Piketty viser til en konveks sammenheng gjennom 1900-tallet (Roine, 2014, s. 22).

Reduksjonen i ulikhet på kort sikt forklarer han med at de rikes formuer blir om lag halvert i forbindelse med 1.verdenskrig og depresjonen på 1920-tallet (Roine, 2014, s. 23). For å forklare hvorfor ulikhetene øker over tid, vektlegger han to teorier:

1. Arbeidsinntektene konsentreres over tid, slik at de med aller høyest arbeidsinntekt distanserer de andre.
2. Kapitalen akkumuleres over tid fordi avkastningen på kapital er større enn veksten i lønningene (Roine, 2014, s. 36).

Piketty er en motvekt til Kuznets teori om hvordan økonomisk vekst på sikt vil resultere i redusert ulikhet og omfordeling.

2.2.2 Effekten gjennom den politiske og sosiopolitiske kanalen

Effekten gjennom den politiske kanalen tar utgangspunkt i at en stor inntektsforskjell vil innebære at medianvelgeren har relativt lavere inntekt. Medianvelgerteoremet diskuterer at samfunn med store ulikheter har en gjennomsnittslønn som er høyere enn medianinntekten (Meltzer og Richard, 1981, s. 914). Dersom en mindre andel av befolkningen blir rikere og en større andel blir fattigere, vil dette på lang sikt føre til et ønske om omfordeling blant majoriteten av befolkningen. I et demokrati vil flertallet dermed sannsynligvis stemme på partier som omfordeler verdier likere.

Større ulikhet gir økt press for beskatning, som i neste omgang har en negativ effekt på investeringer både i realkapital og humankapital. Dette resulterer i et lavere aktivitetsnivå i økonomien og mindre vekst på lang sikt. Studier publisert av Alesina og Rodrik viser at ulikhet i eiendom og inntekter er negativt korrelert med investeringer i fysisk kapital (Alesina og Rodrik, 1994, s. 465). Persson og Tabellini diskuterer på sin side den negative korrelasjonen mellom inntektsulikhet og humankapital (Persson og Tabellini, 1994, s. 617). Denne sammenhengen kan knyttes opp mot medianvelgerteoremet. Studiene konkluderer med at inntektsulikhet er skadelig for vekst, fordi det ofte resulterer i politikk som ikke beskytter eiendomsrett, samt reduserer incentivet til å investere i humankapital. Dette vil igjen bidra til lavere vekst, siden humankapital ansees som en viktig vekstfaktor.

Ulikhet kan også påvirke økonomisk vekt gjennom sosial misnøye. Alesina og Perotti målte sosial uro i form av blant annet antall protester, streiker, vold og kupp. De kom frem til at økt

inntektsulikhet førte til sosial uro, som påvirker den økonomiske veksten negativt (Alesina og Perotti, 1996, s.1). Bénabau på sin side, målte sosial misnøye ved å bruke eiendomsrett, eller mer spesifikt risiko knyttet til investering, grad av lovhåndtering, grad av kontraktsbrudd og omfang av korrupsjon. Han konkluderer med at økt grad av ulikhet forverrer de to nevnte målene over, og er korrelert med lavere investering og vekst (Bénabau, 1996, s. 49).

2.2.3 Effekten av ulikhet gjennom et imperfekt kredittmarked

Kredittmarked påvirker sammenhengen mellom ulikhet og vekst. Kreditt brukes blant annet for å utjevne konsum, samt realisere investeringer. Informasjonen vil i ulik grad være asymmetrisk i forbindelse med transaksjonene som gjennomføres. Med dette mener vi at utlåner ikke alltid har full oversikt eller kjennskap til sentral informasjon. Dette er en vesentlig faktor for å kunne vurdere hvor attraktivt det er å låne ut penger, og skaper usikkerhet i markedet.

Modellen til Banerjee og Duflo illustrerer hvordan asymmetrien i kredittmarkedet og pant påvirker lånesituasjonen (Banerjee og Duflo, 2010, s. 62). Det blir observert en tendens hvor fattige land er informelle utlånere, med svært høye utlånsrenter. Dette gir et mer imperfekt kredittmarked hvor asymmetrien i transaksjonene er en sentral utfordring for fattige land. Dette synspunktet er en av årsakene til teorien om at ulikhet påvirker veksten i fattige land på en annen måte enn i rike land (Galor, 2012, s.26). Rike land løser ofte asymmetrien i kredittmarkedet med å stille pant for lån, noe som gir sikkerhet for lånet ved å for eksempel knytte det til eiendom. Dette resulterer i sikrere informasjonsflyt og lettere gjennomførbare lånetransaksjoner. For fattige land er ikke denne prosedyren like vanlig, siden det som regel ikke strekker til nok ressurser for å kunne stille pant for lån. Galors teori om det imperfekte kredittmarkedet belyser at land med stor ulikhet vil ha flere individer som ekskluderes fra kredittmarkedene, siden de ikke har muligheten til å ta opp lån. Dette vil til syvende og sist også påvirke økonomisk vekst negativt.

3. Data

Det skal i denne delen redegjøres for variabler som har blitt brukt i oppgavens regresjonsanalyse, og hvorfor de er relevante i sammenhengen mellom inntektsulikhet og økonomisk vekst. Dataen til variablene er hentet fra Verdensbanken. Verdensbankens database inneholder samlinger av tidsseriedata og gir omverdenen informasjon om viktige

økonomiske nøkkeltall og statistikk (The World Bank, 2023). Tidsseriedataen vi benytter strekker seg fra 1960 til 2021, med observasjoner for de aktuelle variablene. Dataen tar for seg 81 høyinntektsland og 82 lavinntektsland.

3.1 Variabler

Modellen består av en avhengig og fem uavhengige variabler. Økonomisk vekst er den avhengige variabelen, mens inntektsulikhet er forklaringsvariabelen, samt fem kontrollvariabler. Vi har valgt befolkningsvekst, sparerate, inflasjon, offentlig forbruk og investeringsrate i prosent av BNP som kontrollvariabler.

3.1.1 Avhengig variabel

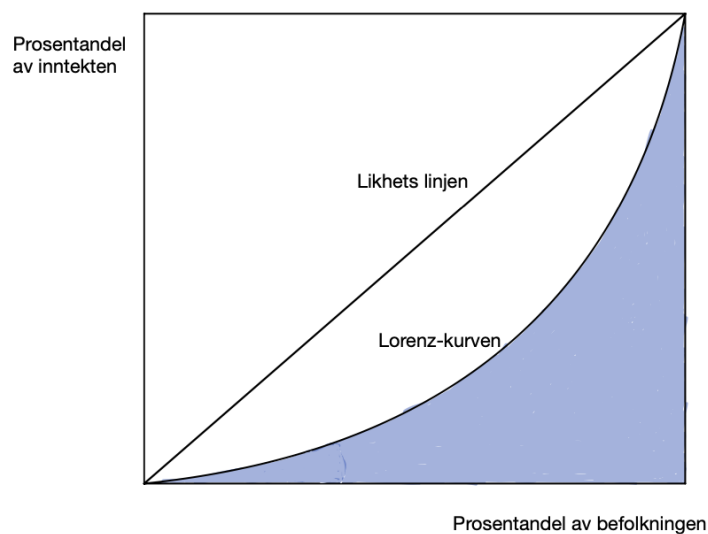
Økonomisk vekst defineres som endring i samlet produksjon til et land, og måles gjennom den årlige prosentvise veksten i BNP per innbygger. BNP er definert som den endelige produksjonen av varer og tjenester produsert av landets økonomi, innenfor landets territorium, av innbyggere og ikke-bosatte, uavhengig av produksjonens fordeling mellom innenlandske og utenlandske verdier (Todaro og Smith, 2020, s. 43). BNP illustrerer dermed den gjennomsnittlige årlige veksten for et land i markedspriser, og på den måten kan vi sammenligne økonomisk vekst på tvers av land.

BNP per innbygger korrigerer samlet produksjon for populasjonsstørrelse. Dermed kan en sammenligne utvikling på tvers av land, uten at landets størrelse har noe innvirkning. Det er det vanligste og mest brukte målet på økonomisk vekst, og dermed var det enkelt å finne data for alle landene i regresjonen. På den andre siden tar målet kun hensyn til størrelsen på befolkningen, men ikke hvordan veksten er fordelt innad i landet. Videre er nødvendigvis ikke BNP per innbygger et godt mål for velstand, ettersom det ikke sier noe om velferd eller lykke for landets innbyggere. I dataen vår måles BNP i konstante 2015-priser i US-dollar. Størrelsen på BNP er summen av brutto verdiskapning for husholdninger, myndigheter og industrier som opererer i økonomien. Verdien beskriver som nevnt alt av innenlandsk produksjon, uavhengig om inntekten tilfaller innenlandsk eller utenlandske bedrifter (The World Bank, 2023). I analysen vår er variabelen angitt som *percapitagdp*.

3.1.2 Forklaringsvariabel

Det finnes ulike måter å beregne inntektsulikhet, hvor vi i denne analysen bruker et lands Gini-koeffisient som måleenhet. Gini-koeffisient viser den relative graden av inntektsulikhet i et land, og kan variere fra 0, som utgjør perfekt likhet, til 1, som tilsier perfekt ulikhet (Todaro og Smith, 2020, s. 226). Vi har skalert Gini-koeffisienten til prosent, og benytter dermed 0 til 100 deretter.

Figur 3: Lorenz-kurven



Gini-koeffisienten tar utgangspunkt i Lorenz-kurven. Denne illustrerer ulikheten til et land grafisk ved å plote de kumulative prosentene av den totale inntekten til et land mot antallet innbyggere.

Gini-koeffisienten måler arealet mellom Lorenz-kurven og en tenkt linje med absolutt likhet, uttrykt som en prosentandel av det maksimale arealet under linjen (The World Bank, 2023). I likningen nedenfor er koeffisienten uttrykt matematisk. Likningen tar utgangspunkt i ulikhetene mellom alle par av inntekter og summerer den totale differansen. Størrelsen normaliseres så ved å dividere på populasjonen kvadrert (n^2), inntektssnittet (μ) og 2, mens y utgjør inntekt. Resultatet vil være en størrelse mellom 0 og 1.

$$G = \frac{1}{2n^2\mu} \sum_j^n \sum_k^m n_j n_k |y_j - y_k|$$

Et land med høy koeffisient, altså at verdien er nær 1, vil ha en høyere grad av inntektsulikhet kontra et land med verdi nær 0. Vi kan dermed bruke koeffisienten til å vurdere i hvilken grad fordelingen av inntekt mellom individer eller husholdninger avviker fra en helt lik fordeling. En kritikk av Gini-koeffisienten er at den ikke sier noe om hvordan inntekten er fordelt innad i de ulike inntektsgruppene. To land med samme Gini-koeffisient kan ha forskjellig fordeling av inntektene (Todaro og Smith, 2015, s. 240). Forklaringsvariabelen blir angitt som *Gini* i analysedelen.

3.1.3 Valg av kontrollvariabler

I følgende avsnitt vil det redegjøres for oppgavens fire kontrollvariabler. Disse representerer sentrale forhold som både påvirker økonomisk vekst og inntektsulikhet. Ved å kontrollere for disse variablene i regresjonsanalysen, får vi i større grad isolert forholdet mellom økonomisk vekst og inntektsulikhet.

Befolkningsvekst

Befolkningsvekst måles som den prosentvise endringen i midtårsbefolkningen fra år $t - 1$ til år t (The World Bank, 2023). Tidligere forskning viser at økonomisk vekst kan knyttes til befolkningsvekst. Solow-modellen med befolkningsvekst predikerer en negativ effekt av økt befolkning på økonomisk vekst (Canarella og Pollard, 2003, s. 91). Andre forskningsresultater argumenter for at befolkningsvekst i spesifikke situasjoner kan ha en positiv effekt (Fengler, 2010). Befolkningsveksten vil være angitt som *popgrowth* videre i regresjonsanalysen.

Sparerate som andel av BNP

Spareraten er beregnet som nasjonalinntekten trukket fra totalt konsum, pluss netto overføringer fra og til utlandet. Brutto sparing viser forskjellen mellom disponibel inntekt og konsum, og måles i prosent av BNP (The World Bank, 2023). Den inkluderes som en kontrollvariabel fordi flere forskningsresultater og økonomiske modeller indikerer en positiv sammenheng mellom spareraten og økonomisk vekst. Easterly og Wetzel (1989) og Fischer (1992) konkluderer med at spareraten har en positiv effekt på økonomisk vekst (Chirwa og Odhiambo, 2016, s. 34). Dataen har blitt hentet fra verdensbankens database. Videre i regresjonsanalysen vil spareraten som prosent av BNP være angitt som *savings*.

Inflasjon

Inflasjon er målt ved konsumprisindeksen (KPI), som illustrerer den årlige prosentvise endringen i kostnaden for gjennomsnittsforbrukeren ved å anskaffe en varekurv til et visst tidspunkt, for eksempel årlig (The World Bank, 2023). Denne variabelen inkluderes fordi den har betydelige makroøkonomiske virkninger. Store variasjoner i prisene kan føre til forsterkede svingninger i økonomien. Dette kommer av at husholdninger og bedrifter blir usikre på sine framtidige inntekter og utgifter, noe som i neste omgang fører til endringer i deres langsiktige økonomiske prioriteringer (Norges Bank, 2019). Inflasjon er betegnet som *inflation* i regresjonsanalysen.

Offentlige utgifter som andel av BNP

Offentlige utgifter omfatter det totale forbruket av varer og tjenester for offentlig sektor og måles som årlig prosent av BNP. Variabelen inkluderer utgifter til nasjonalt forsvar og sikkerhet, men utelater statlige militærutgifter som er en del av statlig kapitaldannelse (The World Bank, 2023). Variabelen er inkludert fordi empiriske studier har gitt motstridende resultater på hvordan offentlige utgifter påvirker økonomisk vekst, og det er dermed interessant å studere effekten den har på dataen vi ser på.

På den ene siden har vi hypoteser om at en økning i offentlige utgifter påvirker økonomisk vekst i negativ forstand (Barro, 2003, s. 231). Andre konkluderer med at offentlige utgifter er positivt knyttet til økonomisk vekst (Ram, 1986, s.191). mens enkelte ikke finner en signifikant sammenheng. Variabelen blir betegnet som *government* i videre regresjon.

Fordeling av land

Tidligere studier illustrerer at det foreligger forskjellige funn for fattige og rike land (Galor, 2012, s. 49). Med dette tatt i betraktning, har vi valgt å gjennomføre to ulike regresjonsanalyser, en for lavinntektsland, og en for høyinntektsland. Dette er for å teste om fattige og rike land har et ulikt forhold mellom ulikhet og vekst.

En tradisjonell måte å definere nivåer av økonomisk utvikling på er inntekt per innbygger. Verdensbanken har et anerkjent system for inntektssammenligninger. Systemet har samlet inn data på 1 216 økonomier med en befolkning på minst 30 000, og rangert økonomiene etter nivåene av BNI per innbygger. Disse økonomiene blir deretter klassifisert som

lavinntektsland, lavere mellominntektsland, høyere mellominntektsland og høyinntektsland (Todaro og Smith, 2020, s. 37).

Siden de ulike kategoriene er basert på BNI per innbygger, er det relevant å diskutere hvorvidt dette er et godt mål på et lands utviklingsnivå. Bruttonasjonalinntekt er inntekt fra innenlandsk produksjon og formue plasseringer i utlandet, som tilfaller de som bor i et land. BNI er altså bruttonasjonalprodukt minus formuesinntekt og lønn til utlandet, netto (The World Bank, 2023). På den ene siden er BNI per innbygger et begrenset mål, siden det ikke tar hensyn til aktiviteter som ikke direkte skaper verdiskapning, noe som er relevant for lavinntektsøkonomier. Økonomier som i stor grad består av selvforsyning og jordbruk, fremfor industri og næringsliv, vil ha et lavere nivå av BNI per innbygger. Videre reflekterer ikke BNI rundt ulikheter i inntektsfordelingen. Dette kan rette tvil på hvor godt målet beskriver utviklingen til et land.

Til tross for at det foreligger nevneverdige begrensninger til BNI per innbygger, så er det en nyttig og enkel indikator å bruke i slike analyser, hvor målet er nært korrelert med andre ikke-monetære mål på livskvalitet og utvikling.

I 2023 er det 28 land som klassifiseres som lavinntektsøkonomier, og 54 land som faller inn under kategorien lavere mellominntektsøkonomier. Den første regresjonen vår består av disse to kategoriene, og vil være en regresjon på 82 land, som har en BNI per innbygger på under \$4,256. Den andre regresjonen vår består av den siste kategorien, nemlig høyinntektsland. Her har vi data på 81 land, som har en BNI per innbygger på \$13 205 eller mer.

3.2 Deskriptiv statistikk

I denne delen av oppgaven skal det redegjøres for viktige egenskaper ved det anvendte datasettet. Dette innebærer følgende mål på sentraltendens og spredning; Gjennomsnitt, standardavvik, minimums- og maksimumsverdi på de ulike variablene. Tabell 1 viser deskriptiv statistikk for hele datasettet, mens Tabell 2 viser deskriptiv statistikk for lavinntekts- og høyinntektslandene hver for seg. Alle observasjoner som mangler data på *Gini* har vi valgt å utelate fra statistikken, da disse heller ikke vil være inkludert i modellen. Denne delen vil i hovedsak kommentere den deskriptive statistikken for hele datasettet, men vil også ta for seg interessante forskjeller i den deskriptive statistikken mellom de to utvalgene.

Tabell 1: Deskriptiv statistikk for hele datasettet

Variabel	Obs	GJ. Snitt	Std.avvik	Min	Maks
<i>Percapitagdp</i>	1196	2.094	3.529	-14.759	23.201
<i>Gini</i>	1211	36.013	8.199	20.7	65.8
<i>Government</i>	1160	17.374	7.186	0.911	136.350
<i>Inflation</i>	1188	10.194	138.660	-4.478	4734.915
<i>Savings</i>	1079	21.755	10.302	-13.984	223.954
<i>Popgrowth</i>	1211	1.34	1.140	-2.258	5.005

Variabelen *Percapitagdp* viser et stort sprik i den årlige vekstraten i BNP per innbygger blant landene. Den høyeste vekstraten ligger på 23.201%, mens den laveste viser en negativ vekstrate på -14.759%. Den gjennomsnittlige vekstraten ligger på 2.094%, med et standardavvik på 3.529. *Gini* viser også en relativt varierende grad av inntektsulikhet, der de med høyest grad av ulikhet har en Gini-koeffisient på 65.8%, mens de med lavest ulikhet har en verdi på 20.7%. Den gjennomsnittlige verdien på Gini-koeffisienten ligger på 36.013%. *Gini* har også et høyere standardavvik enn *Percapitagdp* på 8.199, noe som indikerer større variasjon i grad av inntektsulikhet enn vekst i BNP per innbygger.

Government viser at det laveste nivået av offentlige utgifter ligger på 0.911% av BNP, mens det høyeste nivået er på 136.35% av BNP. Selv om den sistnevnte verdien fremstår som noe urealistisk, kan den høye verdien for eksempel forklares av store offentlige lån. Den gjennomsnittlige verdien på offentlige utgifter ligger på 17.474% av BNP, med et standardavvik på 7.186.

Inflation viser at den høyeste veksten i konsumprisene i disse landene var på 4734.915%. Denne ekstremverdien kommer fra Ukraina i 1993. Den laveste verdien er derimot på -4.478% og viser til Irlands deflasjon under finanskrisen i 2009. Den gjennomsnittlige inflasjonen ligger på 10.194%. Standardavviket ligger på 138.660 og viser til stor spredning i dataen.

Savings gir en maksverdi på 223.954% av BNP. Denne ekstremverdien i spareraten fant sted i Øst-Timor i 2007. Minimumsverdien på -13.984% av BNP er det Elfenbenkysten som

står for i 2008. En negativ sparerate vil si at befolkningen i snitt har konsumert for høyere verdier enn deres disponible inntekt, og må derfor ha tatt opp lån for å finansiere forbruket. Den gjennomsnittlige spareraten er på 21.755% med et standardavvik på 10.302. *Popgrowth* viser at den høyeste årlige befolkningsveksten var 5.005%, mens den laveste viser til en befolkningsnedgang på -2.258%. Gjennomsnittlig befolkningsvekst er lik 1.34% med et standardavvik på 1.14.

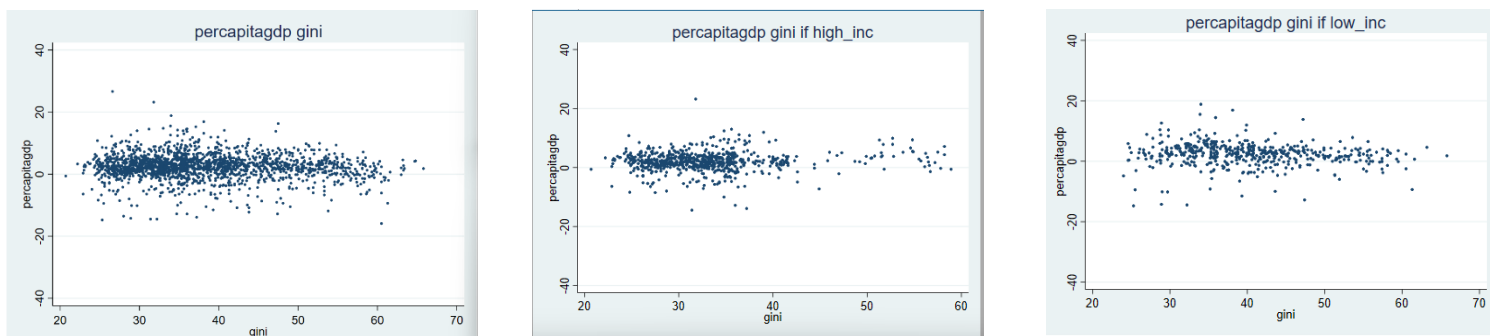
Tabell 2: Deskriptiv statistikk for lavinntektsland og høynntektsland

Obs	Lavinntektsland 498					Høynntektsland 713				
	Variabel	Obs	GJ. Snitt	Std.avvik	Min	Maks	Obs	GJ. Snitt	Std.avvik	Min
Percapitagdp	493	2.262	3.832	-14.759	18.849	703	1.975	3.297	-14.464	23.201
Gini	498	40.25	8.457	24	65.8	713	33.053	6.556	20.7	58.9
Government	453	14.494	9.904	0.911	136.350	707	19.219	3.637	9.674	41.890
Inflation	475	21.375	218.886	-4.476	4734.915	713	2.745	4.324	-4.478	78.310
Savings	411	20.492	14.693	-13.984	223.954	668	22.533	6.102	1.558	41.889
Popgrowth	498	1.924	1.045	-1.152	5.005	713	0.582	0.839	-2.258	3.931
Gini2	498	1691.408	719.161	576	4329.64	713	1135.441	506.498	428.49	3469.21

Høynntektslandene har generelt en lavere grad av spredning og færre ekstremverdier i dataen sammenliknet med lavinntektslandene. Trenden vi ser er at høynntektslandene har en høyere vekstrate i BNP og en lavere Gini-koeffisient i forhold til lavinntektslandene. Datasettet består totalt av flere observasjoner for høynntektsland enn lavinntektsland, med til sammen 1211 observasjoner fra 163 land.

Datasettet inneholder en rekke ekstremverdier. Dette skyldes blant annet at vi bruker et stort datasett med tidsseriedata, som inneholder 1211 observasjoner fra 163 land over et tidsspenn på 61 år. Det er viktig å merke seg at gjennomsnittenes evne til å måle dataens sentraltendens svekkes av ekstremverdiene. Det er også en relativt stor grad av variasjon i dataen, noe som gjør det mer krevende å tyde gjennomgående mønstre i variablene blant disse landene.

Figur 4, 5 og 6: Sammenheng mellom *Percapitagdp* og *Gini*



Alle land

Høyinntektsland

Lavinntektsland

De tre figurene viser henholdsvis sammenhengen for alle landene i analysen, høyinntektslandene og lavinntektslandene. Vi ser tydelige likhetstrekk mellom de tre grafene, der det ikke ser ut til å være noe direkte sammenheng mellom *Gini* og *Percapitagdp*. Uavhengig av nivået på *Gini* ser vi at de fleste observasjonene i spredeplottene har en *Percapitagdp* verdi rundt gjennomsnittet. Selv om det ikke er en tydelig sammenheng i grafene, betyr det nødvendigvis ikke at det ikke eksisterer en sammenheng mellom ulikhet og vekst, ettersom det er mye som ikke kontrolleres for i disse grafene. Det er derimot interessant at høyinntektslandene stort sett har lavere Gini-verdier, mens lavinntektslandene har en spredning i Gini-verdiene. Det sier ingenting om korrelasjonen mellom Gini-koeffisienten og veksten i BNP per innbygger, men viser at landene som befinner seg i kategorien høyinntektsland stort sett har en lav Gini-koeffisient.

4. Metode

I denne oppgaven anvendes det to multiple lineære regresjonsmodeller (MLR), for å undersøke om det er en signifikant sammenheng mellom inntektsulikhet og økonomisk vekst. Modellene består av en avhengig variabel (y), og en rekke uavhengige variabler (x_i). Sistnevnte omfatter en rekke kontrollvariabler, og variabelen av interesse. De brukes for å estimere de relevante populasjonsparameterne gjennom minste kvadraters metode (OLS), og mer spesifikt effekten av én enhets endring i den uavhengige variabelen av interesse på den

avhengige variabelen. I motsetning til enkel lineær regresjon (SLR) er MLR en mer gunstig metode for å finne kausale sammenhenger mellom disse variablene, fordi den lar oss kontrollere for andre faktorer som påvirker den avhengige variabelen. Når det inkluderes flere kontrollvariabler i analysen vil naturligvis en større del av variasjonen i y kunne forklares av modellen (Wooldridge, 2019, s.60).

Minste kvadraters metode sørger for estimater som minimerer summen av kvadrerte residualer. Residualet til observasjon i kan forstås som differansen mellom dens reelle verdi og dens tilpassede verdi. Videre lar den oss undersøke om det er en statistisk signifikant sammenheng mellom inntektsulikhet og økonomisk vekst og om denne eventuelt er positiv eller negativ (Wooldridge, 2019, s.60). Den generelle multiple regresjonsmodellen for populasjonen kan skrives på følgende måte:

$$(i) \quad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k + u$$

I denne likningen er det k uavhengige variabler og $k + 1$ ukjente populasjonsparametere. β_0 er kjent som parameteren for konstantleddet. Den viser verdien på den avhengige variabelen (y) når alt annet er lik null. Parameteren β_k viser den estimerte størrelsen på endringen i y ved én enhets økning i de ulike uavhengige variablene x_k , og representerer dermed helningen til grafen til y . Variabelen u kalles for feilleddet eller forstyrrelsen i sammenhengen vi ser på, og representerer alle andre faktorer enn x_k som påvirker y (Wooldridge, 2019, s.21 og 63).

Ved å minimere summen av de kvadrerte residualer, gir minste kvadraters metode (OLS) oss regresjonsfunksjonen for utvalget, også kalt OLS-regresjonslinjen. Med utgangspunkt i utvalget vårt gir denne estimater eller såkalte tilpassede verdier på de ekte populasjonsparametere fra likning (i).

$$(ii) \quad \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \dots + \hat{\beta}_k x_k$$

Notasjonen med hatter betyr at de predikerte verdiene til variablene i likningen er estimater for populasjonen. Disse estimatene vil naturligvis ha en feilmargin, noe som impliserer at estimatene til en viss grad vil variere for ulike utvalg. Den ekte populasjons regresjonsfunksjonen er ukjent (Wooldridge, 2019, s.28 og 64).

4.1 Forutsetninger for minste kvadraters metode

For å gi objektive og effektive estimater på forholdet mellom den avhengige- og de uavhengige variablene, hviler modellene på en rekke forutsetninger som vi videre skal redegjøre for.

4.1.1 MLR.1 – Linearitet i parametere

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

Linearitet i parameterne er en forutsetning for anvendelsen av OLS-metoden, og innebærer at den partiell deriverte av y med hensyn på hver av regresjonskoeffisientene β er en funksjon av kjente konstanter. Det må også inkluderes et feilledd i likningen, som fanger opp andre faktorer som påvirker den avhengige variabelen og som ikke er inkludert i modellen. Variabler kan også inkluderes ikke lineært, men dette forutsetter at parameterne også er inkludert lineært (Wooldridge, 2019, s.74).

4.1.2 MLR.2 – Tilfeldig utvalg

Det må foreligge et tilfeldig utvalg bestående av n observasjoner: $\{(x_{i1}, x_{i2}, \dots, x_{ik}, y): i = 1, 2, \dots, n\}$. For at utvalgene i analysen vår skal være representative for populasjonen, forutsettes det at valget av utvalg blir gjennomført tilfeldig og at hver observasjon, i , har den samme sannsynligheten for å bli valgt (Wooldridge, 2019, s.74).

4.1.3 MLR.3 – Ingen perfekt kollinearitet

MLR.3 antar to ting. For det første skal ingen av de uavhengige variablene i utvalget, og derfor også i populasjonen, være konstante. Det skal heller ikke være noen eksakt lineært forhold blant de uavhengige variablene. Hvis en uavhengig variabel i modellen er en eksakt lineær kombinasjon av de andre uavhengige variablene, sies det at modellen lider av perfekt kollinearitet, noe som fører til at den ikke kan estimeres av minste kvadraters metode. De uavhengige variablene kan etter denne antakelsen være korrelert, men de kan ikke være perfekt korrelert. På den andre siden må det være noe variasjon mellom variablene. Hvis variansen deres er lik 0 holder heller ikke denne forutsetningen (Wooldridge, 2019, s.74 og 75).

4.1.4 MLR.4 – Null-betinget gjennomsnitt

$$E(u | x_1, x_2, \dots, x_k) = 0$$

Feilleddet u har en forventet verdi lik null, gitt en hvilken som helst verdi av de uavhengige variablene. Dette betyr at u ikke kan være korrelert med disse. Når denne antakelsen holder, sier vi at vi har eksogene uavhengige variabler. Hvis x_j derimot er korrelert med u , sier vi at x_j er en endogen uavhengig variabel. Dessverre vil vi aldri vite med sikkerhet om den gjennomsnittlige verdien til den uobserverte faktoren ikke er relatert til de uavhengige variablene. Likevel er dette en kritisk antakelse for modellens objektivitet.

Modellen antas å være objektiv under antakelsene MLR.1 til MLR.4, noe som innebærer at estimatene fra utvalget korresponderer til populasjonen; $E(\hat{\beta}_j) = \beta_0$ (Wooldridge, 2019, s.76 og 77).

4.1.5 MLR.5 – Homoskedastisitet

$$\text{Var}(u | x) = \sigma^2$$

Denne forutsetningen innebærer at feilbegrepet u har den samme variansen gitt hvilken som helst verdi av de uavhengige variablene. Hvis variansen derimot endres med noen av variablene, holder ikke MLR.5 og modellen viser heteroskedastisitet. Alene er ikke denne forutsetning viktig for modellens objektivitet, men en forutsetning for å estimere standardavvikene. Antakelsene MLR1 til MLR5 er kjent som Gauss-Markov antakelsene (Wooldridge, 2019, s.82 og 83). Under disse er $E(\sigma^2) = \sigma^2$ (Wooldridge, 2019, s.88). Under Gauss-Markov forutsetningene er OLS-estimatoren den beste lineære objektive estimatoren (BLUEs) (Wooldridge, 2019, s.92).

4.1.6 MLR.6 – Normalitetsforutsetningen

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

Normalitetsforutsetningen krever at populasjonens feilledd, u , ikke avhenger av de uavhengige variablene x_1, x_2, \dots, x_k og er normalfordelt med null som gjennomsnitt og en varians σ^2 : $u \sim \text{Normal}(0, \sigma^2)$.

Forutsetningene MLR.1 til MLR.6 kalles for de klassiske lineære modell forutsetningene (CLM). Under disse forutsetningene vil estimatene fra minste kvadraters metode $\hat{\beta}_0, \hat{\beta}_1 \dots \hat{\beta}_k$ være mer effektive enn under Gauss-Markov antakelsene (Wooldridge, 2019, s.106)

Regresjonsanalysen bruker tidsserie-data, noe som medbringer noen implikasjoner for forutsetningene ovenfor. En viktig forutsetning når man jobber med tidsserie-data er at det ikke foreligger seriekorrelasjon. Dette innebærer at feilledet på tidspunkt t er uavhengig av feilledet på tidspunkt $t-1$. Videre forutsettes det at feilledene ($u_t, u_{t+1} \dots u_{t+n}$) i forskjellige perioder ikke er korrelerte, slik at den serielle korrelasjonen er lik 0, altså $\text{Corr}(u_t, u_s | X) = 0$. (Wooldridge, 2019, s.344). Dersom denne forutsetningen ikke holder sier vi at feilledene er auto-korrelerte, fordi de er korrelerte over tid. (Wooldridge, 2019, s.346).

4.2 Passform

Passformen i modellen måles med R^2 . Den måler hvor stor andel av variansen i den avhengige variabelen som kan forklares av de uavhengige variablene i modellen (Wooldridge, 2019, s59). Den regnes ut med den totale kvadratsummen (SST), den antatte kvadratsummen (SSE), og den gjenværende kvadratsummen (SSR). Der SST består av den totale summen av de forklarte og de uforklarte delene i regresjonsmodellen, som også er summen av SSE og SSR. Basert på dette kan vi finne en likning for R^2 .

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST} \rightarrow \frac{SST - SSR}{SST}$$

Verdien på R^2 vil alltid være mellom 0 og 1. Dersom verdien er 0 vil det ikke være noe korrelasjon mellom de avhengige variablene og den uavhengige variabelen. På den andre siden vil en R^2 verdi på 1 si at OLS-estimatoren passer perfekt med dataen. Som regel vil ikke R^2 være 0 eller 1, men heller en verdi mellom disse. Dersom sammenhengen mellom SSR og SST er høy, vil forholdet mellom den uavhengige variabelen og avhengige variablene være lav.

I regresjonen vil vi derimot bruke den justerte R^2 . Denne tar hensyn til at R^2 vil øke for hver kontrollvariabel som legges til i modellen. Dermed vil den justerte R^2 vekte den forklarte variansen opp mot det totale antallet variabler i modellen. Dette betyr at modellen ikke automatisk vil få høyere verdier for passform om flere kontrollvariabler legges til i modellen. Verdien på R^2 øker kun om kontrollvariabelen som legges til forsterker modellens

forklaringskraft. Dersom den svekker modellen, vil verdien på den justerte R^2 avta (Wooldridge, 2019, s.181).

4.3 Hypotesetesting

Hypotesetesting brukes for å trekke slutninger om populasjonsparametere β_j ut ifra utvalgsparametere $\hat{\beta}_j$ under et gitt signifikansnivå. Det er med andre ord en måte å teste om det er sannsynlig at en observert effekt eller sammenheng mellom ulike variabler i et utvalg er reell og signifikant. Dette gjøres ved å formulere en nullhypotese (H_0) og en alternativhypotese (H_A). Mens nullhypotesen beskriver en påstand som antar at det ikke er noen signifikant sammenheng mellom de avhengige- og de uavhengige variablene, beskriver alternativhypotesen det motsatte. Dette innebærer at de to hypotesene er gjensidig utelukkende. En vanlig framgangsmåte for å teste nullhypotesen er gjennom en t-test. Ut ifra denne testen vil enten H_0 forkastes eller beholdes. En forkastet nullhypotese vil gi evidens og støtte for alternativhypotesen. Til tross for dette er det viktig å merke seg at dette ikke impliserer at vi har bevist at alternativhypotesen er sann (Wooldridge, 2019, s.108 og 109).

Den statistiske programvarepakken STATA gjennomfører en t-test for oss ved å gi P-verdier. P-verdien defineres som et sannsynlighetsmål som gir evidens mot H_0 fra utvalget og viser sannsynligheten for å finne en testobservator som er mer eller like ekstrem som den observerte, gitt at nullhypotesen er sann. Testobservatoren er en numerisk verdi som måler hvor mange standardavvik estimatet $\hat{\beta}_j$ avviker fra den anslåtte verdien til populasjonsparameteren β_j . Den følger en t-fordeling; t_{n-k-1} , hvor $n - k - 1$ representerer antall frihetsgrader, der n er lik antallet observasjoner i utvalget og k er antallet uavhengige variabler (Wooldridge, 2019, s. 116). Lavere P-verdi indikerer mer evidens (Anderson et al. 2020, s. 407). Vi forkaster H_0 hvis P-verdien er lavere enn signifikansnivået ($P \leq \alpha$). Signifikansnivået (α) viser den maksimale sannsynligheten vi er villige til å akseptere for å gjøre en type 1 feil, og dermed forkaste en sann nullhypotese. Det mest vanlige signifikansnivået er på 5%, noe som impliserer at vi forkaster H_0 med 95% sikkerhet (Wooldridge, 2019, s.110).

I denne oppgaven skal det testes om det finnes en statistisk signifikant sammenheng mellom inntektsulikhet, målt med Gini-koeffisienten, og økonomisk vekst målt med BNP per innbygger. Til dette formålet vil det være hensiktsmessig å anvende en to-sidet test. Den lar

oss undersøke om den uavhengige variabelen påvirker den avhengige variabelen uten å presisere om denne sammenhengen må være positiv eller negativ.

En t-test lar oss kun teste en variabels signifikans alene (Wooldridge, 2019, s. 109). Siden vi er interessert i å teste de to variablene $Gini$ og $Gini^2$ samtidig, er vi nødt til å gjennomføre en felles hypotesetest (joint hypothesis test). Dette gjennomføres ved å observere hvor stor del av variasjonen i y som forblir uforklart når disse to variablene inkluderes. Med andre ord hvor stor endring som forekommer i summen av de kvadrerte residualene (SSR) når vi går fra en begrenset til en ubegrenset modell. Dette gjennomføres med en F-test (Wooldridge, 2019, s.128). Formelen for testobservatoren i en F-test er gitt som følgende:

$$FS = \frac{(R_{ur}^2 - R_r^2)}{(1 - R_{ur}^2)} * \frac{n - k - 1}{q}$$

Den viser den relative endringen i modellens forklaringskraft målt med R^2 fra den begrensede (R_r^2) modellen til den ubegrensede modellen (R_{ur}^2). Under H_0 og MLR.1-MLR6 følger testobservatoren en F-fordeling ($F \sim F_{q, n-k-1}$) med frihetsgrader lik $q, n - k - 1$, der q er lik antallet restriksjoner (Wooldridge, 2019, s.129 og 130). Den kritiske verdien viser den laveste absoluttverdien til testobservatoren som vil resultere i en forkastet nullhypotese (Anderson et al. 2020, s. 407 og 409). Den korresponderer til signifikansnivået, og q frihetsgrader i teller og $n - k - 1$ frihetsgrader i nevner (Wooldridge, 2019, s.129 og 130). Vi forkaster H_0 og konkluderer at variablene våre er signifikante i felleskap hvis absoluttverdien til testobservatoren er høyere enn den kritiske verdien ($FS > c$).

5. Regresjonsanalyse

I denne delen vil det presenteres resultater fra to ulike multiple lineære regresjonsanalyser. Analysene ble gjennomført i STATA 17.0, og tar utgangspunkt i minste kvadraters metode. De lar oss utforske om det eksisterer en lineær sammenheng mellom vekst i BNP per innbygger og inntektsulikhet målt med Gini-koeffisienten i lavinntektsland og høyinntektsland. Videre forteller de også om sammenhengene er positive eller negative, og om det finnes ulike trender i de to grupperingene av land. Vi har på bakgrunn av Kuznets og Piketty's teorier også valgt å teste om disse sammenhengene er kvadratiske.

I tabellene nedenfor representerer kolonne (1) en enkel lineær regresjon uten kontrollvariabler. For hver kolonne fra (2) til (5), vil hver enkelt av modellens kontrollvariabler legges til en etter en. I kolonne (6) inkluderes det et kvadratisk begrep for Gini-koeffisienten. Denne viser den endelige multiple regresjonsanalysen, der sammenhengen er kontrollert for sparerate, inflasjon, befolkningsvekst og offentlig konsum. Siden all dataen er målt i prosent, vil koeffisientene også tolkes i prosentpoeng, selv om regresjonen som kjøres er en level-level regresjon.

Modell 1 viser sammenhengen mellom vekst i BNP per innbygger og Gini-koeffisienten for utvalget av lavinntektsland. Modell 2 viser den samme sammenhengen for utvalget av høyinntektsland.

5.1 Modell 1: for lavinntektsland

Den første modellen som estimeres er modellen for lavinntektsland. Modellen består av tidsseriedata med observasjoner for alle de aktuelle variablene fra 1960 til 2021 fra 82 lavinntektsland. I utgangspunktet ville antallet observasjoner synke fra kolonne (1) til (6). Dette skyldes at alle observasjonene som bortfaller ikke eksisterer i datasettet for alle kontrollvariablene som er inkludert i analysen. Dermed har vi valgt å holde antallet observasjoner konstant gjennom analysen og fjernet alle observasjoner som ikke eksisterer for alle kontrollvariablene. Hovedgrunnlaget for dette er at de bortfallene observasjonene kan ha en innvirkning på verdien til *Gini* sin regresjonskoeffisient. Siden vi i hovedsak er interessert i å analysere effekten av kontrollvariablene på *Gini* sin regresjonskoeffisient, vil de bortfallene observasjonene kunne resultere i feilaktige tolkninger av *Gini*.

Følgende modeller estimeres:

$$(1) \text{Percapitagdp}_{Low} = \beta_0 + \beta_1 \text{Gini}_{Low}$$

$$(2) \text{Percapitagdp}_{Low} = \beta_0 + \beta_1 \text{Gini}_{Low} + \beta_2 \text{Savings}$$

$$(3) \text{Percapitagdp}_{Low} = \beta_0 + \beta_1 \text{Gini}_{Low} + \beta_2 \text{Savings}_{Low} + \beta_3 \text{Inflation}_{Low}$$

$$(4) \text{Percapitagdp}_{Low} = \beta_0 + \beta_1 \text{Gini}_{Low} + \beta_2 \text{Savings}_{Low} + \beta_3 \text{Inflation}_{Low} + \beta_4 \text{Popgrowth}_{Low}$$

$$(5) \text{Percapitagdp}_{Low} = \beta_0 + \beta_1 \text{Gini}_{Low} + \beta_2 \text{Savings}_{Low} + \beta_3 \text{Inflation}_{Low} +$$

$$\beta_4 \text{Popgrowth}_{Low} + \beta_5 \text{Government}_{Low}$$

$$(6) \text{Percapitagdp}_{Low} = \beta_0 + \beta_1 \text{Gini}_{Low} + \beta_2 \text{Savings}_{Low} + \beta_3 \text{Inflation}_{Low} +$$

$$\beta_4 \text{Popgrowth}_{Low} + \beta_5 \text{Government}_{Low} + \beta_6 \text{Creditmrkts}_{Low} + \beta_7 \text{Gini}_{Low}^2$$

Tabell 3: Regresjonen for lavinntektsland

Variabler	(1)	(2)	(3)	(4)	(5)	(6)
	<i>Percapitagdp_{Low}</i>	<i>Percapitagdp_{Low}</i>	<i>Percapitagdp_{Low}</i>	<i>Percapitagdp_{Low}</i>	<i>Percapitagdp_{Low}</i>	<i>Percapitagdp_{Low}</i>
<i>Gini_{Low}</i>	-0.067** (0.022)	-0.061** (0.022)	-0.061** (0.021)	-0.054* (0.021)	-0.058** (0.021)	0.497** (0.184)
<i>Savings_{Low}</i>		0.035** (0.013)	0.037** (0.121)	0.036** (0.012)	0.053*** (0.013)	0.051*** (0.013)
<i>Inflation_{Low}</i>			-0.049*** (0.009)	-0.051*** (0.009)	-0.052*** (0.009)	-0.053*** (0.008)
<i>Popgrowth_{Low}</i>				-0.232 (0.181)	-0.255 (0.179)	-0.371* (0.181)
<i>Government_{Low}</i>					-0.075** (0.024)	-0.063** (0.024)
<i>Gini_{Low}²</i>						-0.006** (0.002)
<i>_cons</i>	5.116*** (0.878)	4.097*** (0.941)	4.573*** (0.911)	4.749*** (0.918)	5.625*** (0.951)	-5.438 (3.777)
<i>N</i>	397	397	397	397	397	397
<i>R²</i>	0.024	0.044	0.119	0.122	0.144	0.163
<i>Adj. R²</i>	0.022	0.039	0.112	0.114	0.133	0.150

Standardfeil i parentes

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Modell (1) til (6) estimeres i *Tabell 3*. Verdiene i hver kolonne viser til de estimerte regresjonskoeffisientene ($\hat{\beta}_j$) og deres standardfeil i parentes. Koeffisientene er merket med stjerner (*) som indikerer variabelens signifikansnivå. En stjerne betyr at variabelen er statistisk signifikant under et signifikansnivå på 5%, to stjerner svarer til et signifikansnivå på 1% og tre stjerner svarer til et signifikansnivå på 0.1%. Modellenes tilpassede R^2 , R^2 og antallet observasjoner vises nest nederst i tabellen. R^2 inkluderes fordi den er nødvendig for å senere gjennomføre en felles hypotesetest (joint hypothesis test). Ellers forholder vi oss til den tilpassede R^2 . I kolonne (6) viser denne at den endelige modellen vår forklarer 15% av

variasjonen i vekst i BNP per innbygger blant lavinntektslandene. Samtidig ser vi at forklaringskraften øker med 12.8 prosentpoeng fra kolonne (1) til (6), som følge av at kontrollvariablene og det kvadratiske begrepet inkluderes. Det er likevel stor variasjon i hvor stor økning hver enkelt kontrollvariabel bidrar med.

Fra kolonne (1) til (5) ser vi en negativ lineær signifikant sammenheng mellom Gini-koeffisienten og økonomisk vekst. Når det kvadrerte begrepet av *Gini* inkluderes i kolonne (6) observerer vi et signifikant kvadratisk forhold. Siden $\beta_1 > 0$ og $\beta_7 < 0$ er denne sammenhengen konkav. Regresjonskoeffisienten til $Gini_{Low}$ øker fra -0.067 i modell (1) til -0.058 i modell (5). Den negative effekten dempes altså noe når kontrollvariablene inkluderes. Dette indikerer at noen av disse i utgangspunktet både kan være korrelert med $Gini_{Low}$ og $Percapitagdp_{Low}$. Disse effektene isoleres ut når kontrollvariablene inkluderes i modellen og ikke lenger inngår i feilleddet u . Et unntak for dette er variabelen $Government_{Low}$ som snarere forsterker den negative effekten av $Gini_{Low}$. En mulig forklaring på dette er at denne kontrollvariabelen ikke fanger innvirkningen av andre variabler som inngår i feilbegrepet u og som både kan være korrelert med denne og den avhengige variabelen.

Kolonne (6) gir oss følgende regresjonslikning:

$$Percapitagdp_{Low} = -5.438 + 0.497Gini_{Low} + 0.051Savings_{Low} - 0.053Inflation_{Low} - 0.371Popgrowth_{Low} - 0.063Government_{Low} - 0.006Gini_{Low}^2$$

Likningens konstantledd viser at den årlige veksten i BNP per innbygger i de utvalgte lavinntektslandene estimeres å ligge på -5.438% , gitt at alle de uavhengige variablene er lik null. Dette impliserer altså den urealistiske antakelsen at lavinntektslandet ikke har noen befolkningsvekst, sparerate og inflasjon lik null, ingen offentlig bruk på varer og tjenester, og ikke minst en Gini-koeffisient lik null, og dermed et samfunn med perfekt likhet i inntektsulikhet. Konstantleddet er ikke statistisk signifikant noe som innebærer at vi ikke kan trekke slutningen om at denne sammenhengen eksisterer i populasjonen.

Videre estimerer modellen at veksten vil øke ved en ett prosentpoengs økning i Gini-koeffisienten i avtakende grad, alt annet likt. Etter hvert vil effekten gå over til av være

negativ. Den marginale effekten kan illustreres analytisk ved å derivere regresjons likningen med hensyn på $Gini_{Low}$:

$$\begin{aligned}\frac{\partial Percapitagdp_{Low}}{\partial Gini_{Low}} &= 0.497 - 2 \cdot 0.006 Gini_{Low} \\ &= 0.493 - 0.012 Gini_{Low}\end{aligned}$$

Som illustrert over avhenger den marginale effekten av en økning i Gini-koeffisienten på ett prosentpoeng av hvor stor grad av skjevhet i inntektsfordelingen det i utgangspunktet er i lavinntektslandene. Det er derfor interessant regne ut toppunktet på denne kvadratiske sammenhengen. Dette gjøres ved å sette uttrykket over lik null og løse ut $Gini_{Low}$:

$$0.493 - 0.012 Gini_{Low} = 0$$

$$-0.012 Gini_{Low} = -0.493$$

$$Gini_{Low}^* = \frac{-0.493}{-0.012}$$

$$Gini_{Low}^* = 41.08$$

Ut ifra utregningene ser vi at toppunktet til Gini-koeffisienten er lik 41.08. Dette betyr at den økonomiske veksten i lavinntektslandene estimeres å øke i avtakende grad ved økt inntektsulikhet fram til punktet hvor Gini-koeffisienten er lik 41.08. Hvis skjevheten i inntektsfordelingen øker utover dette, estimeres veksten å synke. Datasettet viser at blant annet Argentina, Bolivia, Bulgaria, Haiti og Kenya hadde en Gini-koeffisient tilnærmet dette nivået i 2021. *Tabell 1* viser at den gjennomsnittlige Gini-koeffisienten blant lavinntektslandene ligger på 40.25%, noe som utfra regresjonslikningen innebærer en vekst på 10.28 prosentpoeng årlig.

Den første kontrollvariabelen $Savings_{Low}$ viser en positiv sammenheng mellom sparerate og økonomisk vekst. Mer spesifikt estimeres den årlige veksten i BNP per innbygger å øke med 0.051 prosentpoeng ved en økning i spareraten som andel av BNP på ett prosentpoeng, alt annet likt.

Den andre kontrollvariabelen $Inflation_{Low}$ viser en negativ sammenheng mellom inflasjon og økonomisk vekst. For hver årlige prosentpoengs økning i konsumprisene estimeres den

årlige veksten i BNP per innbygger å falle med 0.053 prosentpoeng, alt annet likt. $Popgrowth_{Low}$ viser en sterkere negativ sammenheng mellom befolkningsvekst og økonomisk vekst. Den årlige økonomiske veksten forventes falle 0.371 prosentpoeng ved en økning i den årlige befolkningsveksten på ett prosentpoeng, alt annet likt. Den siste kontrollvariabelen vår, $Government_{Low}$, viser en negativ sammenheng mellom offentlig konsum og økonomisk vekst. For hvert prosentpoeng offentlige utgifter og investeringer øker som andel av BNP, estimeres den årlige veksten i BNP å falle med 0.063 prosentpoeng, alt annet likt.

Regresjonsmodellen for lavinntektsland har så langt illustrert en konkav kvadratisk sammenheng mellom Gini-koeffisienten og årlig vekst i BNP per innbygger. STATA gjennomfører en t-test og gir oss en P-verdi som viser at denne sammenhengen er signifikant under et 1% signifikansnivå (indikert med ** i Tabell 3). Videre gjennomføres det en felles hypotesetest for å teste de to variablene $Gini_{Low}$ og $Gini_{Low}^2$ samtidig. Vår ubegrensede modell er gitt som modell (6) mens den begrensede er lik den samme uten $Gini_{Low}$ og $Gini_{Low}^2$:

Begrenset modell: $Percapitagdp_{Low} = \beta_0 + \beta_1 Savings_{Low} + \beta_2 Inflation_{Low} + \beta_3 Popgrowth_{Low} + \beta_4 Government_{Low} + \beta_5 Creditmrkts_{Low}$

Ubegrenset modell: $Percapitagdp_{Low} = \beta_0 + \beta_1 Gini_{Low} + \beta_2 Savings_{Low} + \beta_3 Inflation_{Low} + \beta_4 Popgrowth_{Low} + \beta_5 Government_{Low} + \beta_6 Creditmrkts_{Low} + \beta_7 Gini_{Low}^2$

Vi gjennomfører hypotesetesten under et 5% signifikansnivå. Nullhypotesen vår sier at $Gini_{Low}$ og $Gini_{Low}^2$ ikke har noen felles signifikant effekt på økonomisk vekst blant lavinntektslandene, mens alternativhypotesen sier det motsatte. De kan formuleres som følgende:

$$H_0: \beta_{Gini} = 0, \beta_{Gini^2} = 0$$

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

Testobservatoren er for F-testen er gitt som følgende: $FS = \frac{(R_{un}^2 - R_r^2)}{(1 - R_{un}^2)} * \frac{n-k-1}{q} = \frac{(0.15 - 0.127)}{(1 - 0.15)} * \frac{397-7-1}{2} = 5.26$. Vi henter R^2 til den begrensede modellen (R_r^2) fra STATA ved å kjøre en regresjonsanalyse. Under H_0 og MLR.1-MLR6 følger testobservatoren en F-fordeling ($F \sim F_{q, n-k-1}$) med frihetsgrader lik $q, n - k - 1$ (Wooldridge, 2019, s.130). Den kritiske

verdien til en F-fordeling med $q = 2$ frihetsgrader i teller, $n - k - 1 = 397 - 7 - 1 = 389$ frihetsgrader i nevner og et signifikansnivå på 5% ($\alpha = 0.05$) er lik 3. Siden absoluttverdien til testobservatoren er større enn den kritiske verdien ($5.26 > 3$) forkaster vi nullhypotesen og konkluderer med 95% sikkerhet at de to variablene $Gini_{Low}$ og $Gini_{Low}^2$ har en felles signifikant effekt på økonomisk vekst i lavinntektslandene.

5.2 Modell 2: for høyinntektsland

Den andre modellen som estimeres er modellen for høyinntektsland. I likhet med den første modellen består den av tidsseriedata med observasjoner for alle de aktuelle variablene fra 1960 til 2021 fra 81 høyinntektsland. Vi har naturligvis her også valgt å holde antallet observasjoner konstant gjennom analysen og ekskludere alle observasjoner som ikke eksisterer for alle de uavhengige variablene. Verdensbanken har mer omfattende statistikk på høyinntektslandene med 661 observasjoner mot 397 for høyinntektslandene. Følgende modeller estimeres:

$$(1) \text{Percapitagdp}_{High} = \beta_0 + \beta_1 Gini_{High}$$

$$(2) \text{Percapitagdp}_{High} = \beta_0 + \beta_1 Gini_{High} + \beta_2 Savings_{High}$$

$$(3) \text{Percapitagdp}_{High} = \beta_0 + \beta_1 Gini_{High} + \beta_2 Savings_{High} + \beta_3 Inflation_{High}$$

$$(4) \text{Percapitagdp}_{High} = \beta_0 + \beta_1 Gini_{High} + \beta_2 Savings_{High} + \beta_3 Inflation_{High} + \beta_4 Popgrowth_{High}$$

$$(5) \text{Percapitagdp}_{High} = \beta_0 + \beta_1 Gini_{High} + \beta_2 Savings_{High} + \beta_3 Inflation_{High} + \beta_4 Popgrowth_{High} + \beta_5 Government_{High}$$

$$(6) \text{Percapitagdp}_{High} = \beta_0 + \beta_1 Gini_{High} + \beta_2 Savings_{High} + \beta_3 Inflation_{High} + \beta_4 Popgrowth_{High} + \beta_5 Government_{High} + \beta_6 Creditmrkt_{High} + \beta_7 Gini_{High}^2$$

Tabell 4: Regresjonen for høyinntektsland

Variabler	(1)	(2)	(3)	(4)	(5)	(6)
	<i>Percapitagdp_{High}</i>	<i>Percapitagdp_{High}</i>	<i>Percapitagdp_{High}</i>	<i>Percapitagdp_{High}</i>	<i>Percapitagdp_{High}</i>	<i>Percapitagdp_{High}</i>
<i>Gini_{High}¹</i>	0.047*	0.067***	0.062***	0.096***	0.008	-0.327*
	(0.019)	(0.019)	(0.019)	(0.02)	(0.024)	(0.141)
<i>Savings_{High}</i>		0.096***	0.099***	0.124***	0.093***	0.081***
		(0.019)	(0.021)	(0.021)	(0.021)	(0.022)
<i>Inflation_{High}</i>			0.103*	0.095*	0.084	0.066
			(0.046)	0.046	(0.045)	(0.045)
<i>Popgrowth_{High}</i>				-0.794***	-0.739***	-0.809***
				(0.158)	(0.154)	(0.156)
<i>Government_{High}</i>					-0.253***	-0.268***
					(0.042)	(0.042)
<i>Gini_{High}²</i>						0.004*
						(0.002)
_cons	0.426	-2.483**	-2.645**	-3.862***	4.622**	11.45***
	(0.633)	(0.888)	(0.889)	(0.905)	(1.651)	(3.247)
<i>N</i>	661	661	661	661	661	661
<i>R</i> ²	0.009	0.039	0.046	0.082	0.131	0.139
Adj. <i>R</i> ²	0.008	0.037	0.042	0.077	0.125	0.131

Standardfeil i parentes

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Modell (1) til (6) estimeres i *Tabell 4*. I kolonne (6) viser den tilpassede R^2 at den endelige modellen vår forklarer 13.1% av variasjonen i vekst i BNP per innbygger blant høyinntektslandene. Denne har en 1.9 prosentpoengs lavere forklaringskraft enn den endelige modellen for lavinntektsland, til tross for at den inneholder betydelig flere observasjoner. Den øker med 12.3 prosentpoeng fra kolonne (1) til (6).

Til forskjell fra lav inntektslandene viser dataen fra høyinntektslandene en positiv lineær sammenheng mellom Gini-koeffisienten og økonomisk vekst per innbygger fra kolonne (1) til (5), og et positivt signifikant kvadratisk forhold i kolonne (6). Siden $\beta_1 < 0$ og $\beta_7 > 0$ er denne sammenhengen konveks.

Regresjonskoeffisienten øker fra 0.047 til 0.096 fra kolonne (1) til (4). Med andre ord forsterkes den positive effekten av Gini-koeffisienten når variablene *Savings_{High}*, *Inflation_{High}* og *Popgrowth_{High}* kontrolleres for. Dette tyder på at disse kontrollvariablene ikke isolerer ut effekten til Gini-koeffisienten, men at de snarere forsterker effekten. Her er det også mulig at det finnes ukjente variabler i feilbegrepet som både er korrelert med den avhengige variabelen og de uavhengige variablene.

Regresjonskoeffisienten synker derimot fra 0.096 til 0.008 fra kolonne (4) til (5) når variabelen $Government_{High}$ inkluderes. Den går også fra å være signifikant under et 0.1% signifikansnivå til å ikke lenger være statistisk signifikant. Dette betyr at vi ikke lenger kan trekke slutningen om at denne sammenhengen eksisterer i populasjonen, og at $Government_{High}$ isolerer ut store deler av $Gini_{High}$ sin effekt. Dermed er nærliggende å tro at både Gini-koeffisienten og den årlige veksten i BNP blant høyinntektslandene er korrelert med $Government_{High}$ og dermed egentlig i stor grad kan forklares av finanspolitikken deres. Når det kvadrerte begrepet av $Gini$ inkluderes i kolonne (6) blir $Gini_{High}$ statistisk signifikant igjen. En mulig forklaring på dette er at det kvadratiske begrepet fanger en større andel av variasjonen i den uavhengige variabelen. Likevel må vi være varsomme når vi tolker endringen i signifikansen.

Kolonne (6) gir oss følgende regresjonslikning:

$$\begin{aligned} Percapitagdp_{High} = & 11.45 - 0.347Gini_{High} + 0.081Savings_{High} \\ & + 0.066Inflation_{High} - 0.809Popgrowth_{High} - 0.268Government_{High} + 0.004Gini_{High}^2 \end{aligned}$$

Konstantleddet viser at den årlige veksten i BNP per innbygger i de utvalgte høyinntektslandene forventes å ligge på 11.45%, når alle andre variabler er lik null. Dette er høyere enn det som ble estimert for lavinntektslandene. Likevel er det som tidligere nevnt urealistisk, og sjeldent observert at alle andre variabler er lik null. Modellen estimerer at veksten vil synke ved et prosentpoengs økning i Gini-koeffisienten, alt annet likt, men i avtakende grad. Etter hvert vil effekten gå over til å være positiv. Vi illustrerer den marginale effekten av en økning i Gini-koeffisienten på ett prosentpoeng:

$$\begin{aligned} \frac{\partial Percapitagdp_{High}}{\partial Gini_{High}} &= -0.347 + 2 \cdot 0.004Gini_{high} \\ &= -0.347 + 0.008Gini_{High} \end{aligned}$$

Her vil også den marginale effekten avhenge av størrelsen på Gini-koeffisienten i høyinntektslandene. Videre illustreres bunnpunktet:

$$-0.347 + 0.008Gini_{High} = 0$$

$$-0.008Gini_{High} = -0.347$$

$$Gini_{High}^* = \frac{-0.347}{-0.008}$$

$$Gini_{High}^* = 43.38$$

Vi ser at bunnpunktet til Gini-koeffisienten er lik 43.38. Dette betyr at den økonomiske veksten i høyinntektslandene estimeres å synke i avtakende grad ved økt inntektsulikhet fram til dette punktet. Hvis den øker utover dette, estimeres veksten å øke. Datasettet viser at blant annet Madagaskar og Paraguay hadde en Gini-koeffisient tilnærmet dette nivået i 2021. *Tabell 2* viser at den gjennomsnittlige Gini-koeffisienten blant høyinntektslandene ligger på 33.053, noe som utfra regresjonslikningen resulterer i en vekst på 4,35 prosentpoeng årlig.

$$Per\text{capita}gdp_{High} = 11.45 - 0.347Gini_{High} + 0.081Savings_{High}$$

$$+0.066Inflation_{High} - 0.809Popgrowth_{High} - 0.268Government_{High} + 0.004Gini_{High}^2$$

$Savings_{High}$ viser høyinntektslandenes årlige vekst i BNP forventes å øke med 0.081 prosentpoeng ved en økning i spareraten som andel av BNP på ett prosentpoeng, alt annet likt. $Inflation_{High}$ viser at den årlige veksten i BNP per innbygger estimeres å øke med 0.066 prosentpoeng, for hvert prosentpoengs økning i inflasjonen, alt annet likt. Denne variabelen ikke lenger statistisk signifikant etter at variabelen $Government_{High}$ ble inkludert i kolonne (5), og vi kan dermed ikke trekke slutninger om dette forholdet.

Regresjonslikningen viser videre en sterkere negativ sammenheng mellom befolkningsvekst og økonomisk vekst. Den årlige økonomiske veksten forventes å falle med 0.809 prosentpoeng ved ett prosentpoengs økning i den årlige befolkningsveksten, alt annet likt. Den siste kontrollvariabelen vår, $Government_{Low}$, viser en negativ sammenheng mellom offentlig konsum og økonomisk vekst. For hvert prosentpoengs økning i offentlige utgifter og investeringer som andel av BNP estimeres den årlige veksten i BNP å falle med 0.268 poeng, alt annet likt.

For å oppsummere, kan vi se en generell lik trend i forholdet mellom økonomisk vekst og offentlige utgifter, befolkningsvekst og sparing mellom de to grupperingene av land. Gini-

koeffisienten og inflasjon viser derimot et avvik mellom landene. Fra STATA får vi en P-verdi som viser at sammenhengen mellom Gini-koeffisienten og økonomisk vekst både er signifikant lineært og kvadrert under et 5% signifikansnivå (indikert med ** i *Tabell 4*). Videre gjennomfører vi her også en felles hypotesetest gjennom en F-test for å se om $Gini_{Low}$ og $Gini_{Low}^2$ har en felles signifikant påvirkning.

Begrenset modell: $Percapitagdp_{High} = \beta_0 + \beta_1 Savings_{High} + \beta_2 Inflation_{High} + \beta_3 Popgrowth_{High} + \beta_4 Government_{High} + \beta_5 Creditmrkts_{High}$

Ubegrenset modell: $Percapitagdp_{High} = \beta_0 + \beta_1 Gini_{High} + \beta_2 Savings_{High} + \beta_3 Inflation_{High} + \beta_4 Popgrowth_{High} + \beta_5 Government_{High} + \beta_6 Creditmrkts_{Low} + \beta_7 Gini_{High}^2$

Nullhypotesen vår sier at $Gini_{High}$ og $Gini_{High}^2$ ikke har noen felles signifikant effekt på økonomisk vekst blant høyinntektslandene, mens alternativhypotesen sier det motsatte:

$$H_0: \beta_{Gini} = 0, \beta_{Gini^2} = 0$$

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

Testobservatoren er for F-testen er gitt som følgende: $FS = \frac{(R_{un}^2 - R_r^2)}{(1 - R_{un}^2)} * \frac{n - k - 1}{q} = \frac{(0.139 - 0.131)}{(1 - 0.139)} * \frac{661 - 7 - 1}{2} = 3.03$. Den kritiske verdien til en F-fordeling med $q = 2$ frihetsgrader i teller, $n - k - 1 = 661 - 7 - 1 = 653$ frihetsgrader i nevner og et signifikansnivå på 5% ($\alpha = 0.05$) er lik 3. Siden absoluttverdien til testobservatoren på marginen er større enn den kritiske verdien ($3.03 > 3$) forkaster vi nullhypotesen og konkluder med 95% sikkerhet at de to variablene $Gini_{High}$ og $Gini_{High}^2$ har en felles signifikant effekt på økonomisk vekst i høyinntektslandene.

5.3 Modell 3: Interaksjonsbegrep

I kommende del inkluderes det et interaksjonsbegrep i en generell regresjonsmodell for begge grupperingene av land. Et interaksjonsbegrep defineres som en uavhengig variabel som er et produkt av to uavhengige variabler (Wooldridge, 2019, s.762). I modellen benytter vi to interaksjonsbegrep. Den ene er produktet av den kvadratiske variabelen $Gini^2$ og variabelen $Highinc$, angitt som $Gini^2 \cdot Highinc$. Den andre er produktet av $Gini$ og $Highinc$, angitt som $Gini \cdot Highinc$ i *Tabell 5*. Disse lar oss analysere samspillet mellom variablene, og om

effekten av den ene avhenger av verdien av den andre. *Highinc* er en dummyvariabel som har verdien en for høyinntektsland, og null for lavinntektslandene.

Tabell 5: Modell med interaksjonsbegrep

Variabler	(1) <i>Percapitagdp</i>
<i>Gini</i>	0.197 (0.109)
<i>Highinc</i>	6.122 (3.564)
<i>Gini · highinc</i>	-0.467* (0.184)
<i>Gini</i> ²	-0.003* (0.001)
<i>Gini</i> ² · <i>highinc</i>	0.007** (0.002)
<i>Savings</i>	0.083*** (0.009)
<i>Inflation</i>	-0.002* (0.001)
<i>Popgrowth</i>	-0.679*** (0.088)
<i>Government</i>	-0.159*** (0.018)
_cons	1.509 (2.320)
<i>N</i>	1610
Adj. <i>R</i> ²	0.147

Standardfeil i parentes
* *p* < 0.05 ** *p* < 0.01 *** *p* < 0.001

Kolonne (1) gir oss følgende regresjonslikning:

$$\begin{aligned}
 \text{Percapitagdp} = & 1.509 + 0.197Gini + 6.122Highinc \\
 & -0.003Gini^2 - 0.467Gini \cdot Highinc + 0.007Gini^2 \cdot Highinc + 0.083Savings \\
 & -0.002Inflation - 0.679Popgrowth - 0.159Government
 \end{aligned}$$

Videre vil ikke regresjonskoeffisientene til de andre uavhengige variablene kommenteres, og fokuset vil kun ligge på interaksjonsbegrepet og *Gini*. Vi illustrerer den marginale effekten på økonomisk vekst ved en økt Gini-koeffisient analytisk:

$$\begin{aligned}\frac{\partial \text{Percapitagdp}}{\partial Gini} &= 0.197 + 2 \cdot -0.003Gini - 0.467Highinc + 2 \cdot 0.007Gini \cdot Highinc \\ &= 0.197 - 0.006Gini - 0.467Highinc + 0.014Gini \cdot Highinc\end{aligned}$$

Koeffisienten til det første interaksjonsbegrepet av *Gini* er negativt, mens den andre av *Gini*² er positiv. Dette som indikerer at samspillet mellom begge variablene og *Highinc* bidrar til å endre sammenhengen. Dette kan illustreres analytisk ved å sette *Highinc* lik en:

$$\begin{aligned}\frac{\partial \text{Percapitagdp}}{\partial Gini} &= 0.197 - 0.006Gini - 0.467(1) + 0.014Gini \cdot (1) \\ &= -0.27 + 0.000084Gini\end{aligned}$$

Vi ser nå på effekten for høyinntektsland i forhold til referansekategorien lavinntektsland. Her fremkommer det igjen en konveks sammenheng for høyinntektsland, som likner på funnene fra modell (6) for høyinntektsland. Hvis vi derimot setter *Highinc* lik null, vil begge de to siste leddene i likningen bortfalle:

$$\begin{aligned}\frac{\partial \text{Percapitagdp}}{\partial Gini} &= 0.197 - 0.006Gini - 0.467(0) + 0.014Gini \cdot (0) \\ &= 0.197 - 0.006Gini\end{aligned}$$

Dette gir en konkav sammenheng for lavinntektslandene i forhold til høyinntektslandene, noe som også ligner funnene fra modell (6) for lavinntektsland. For å oppsummere framstår det som at funnene fra Modell 3 er i tråd med flere av funnene både fra Modell 2 for høyinntektsland, og Modell 1 for lavinntektsland. Selv om interaksjonsbegrepet *Gini* · *Highinc* er signifikant under et 5% signifikansnivå og *Gini*² · *Highinc* og *Gini*² under et 1% signifikansnivå, er verken *Gini* eller *Highinc* statistisk signifikante. Dette innebærer at vi ikke kan trekke slutninger til populasjonen rundt forholdet vi har diskutert i denne delen.

6. Gjennomgang av forutsetningene for minste kvadraters metode

Forutsetningene for OLS-regresjon ble presentert i del 3.1. Denne delen kommer til å inneholde tester og utregninger for å se om det er sannsynlig at antakelsene holder for vår modell.

6.1.1 MLR.1 - Linearitet

Som forklart tidligere må alle parameterne være lineære for å benytte seg av minste kvadraters metode. Ikke-lineære parametere kan føre til unødvendige feilberegninger i estimeringen av populasjonen. Variablenes linearitet kan testes ved å inkludere variabelens kvadratiske termer. En slik utvidelse gjør det mulig å undersøke om effekten av x_K endrer seg når x_K øker. Dersom det skjer, vil det være en ikke lineær sammenheng mellom den avhengige og den uavhengige variabelen (Ringdal og Wiborg, 2017). I tabell 6 og 7 er de kvadrerte termene for alle de uavhengige variablene inkludert, og er benevnt som *gini*² *savings*² og så videre. Vi kan nå teste om variablene er lineære eller ikke ved å formulere følgende null- og alternativhypotese:

H_0 : Parameterne er ikke-lineære

H_A : Parameterne er lineære

$H_0: \beta_1 = 0, \beta_2 = 0, \beta_3 = 0, \beta_4 = 0, \beta_5 = 0$

$H_A: \beta_1 \neq 0, \beta_2 \neq 0, \beta_3 \neq 0, \beta_4 \neq 0, \beta_5 \neq 0$

Ved å inkludere de kvadrerte termene får vi følgende resultat:

Tabell 6 og 7: Linearitet i variablene

Utvalg for lavinntektsland:

	(1) Percapitagdp _{Low}
Gini _{Low}	0.400* (0.188)
Savings _{Low}	0.039 (0.023)
Inflation _{Low}	-0.104** (0.003)
Popgrowth _{Low}	0.035 (0.499)
Government _{Low}	-0.014 (0.065)
Gini ² _{Low}	-0.005* (0.002)
Savings ² _{Low}	0.0002 (0.0002)
Inflation ² _{Low}	0.0002* (0.00006)
Popgrowth ² _{Low}	-0.103 (0.140)
Government ² _{Low}	-0.001 (0.001)
_cons	-3.565 (13.883)
N	1610
Justert R ²	0.1611

Standardfeil i parentes

* p < 0.05, ** p < 0.01, *** p < 0.001

Utvalg for høyinntektsland:

	(1) Percapitagdp _{High}
Gini _{High}	-0.431** (0.138)
Savings _{High}	0.420*** (0.087)
Inflation _{High}	0.292** (0.084)
Popgrowth _{High}	-1.142*** (0.197)
Government _{High}	-0.399 (0.308)
Gini ² _{High}	-0.005** (0.002)
Savings ² _{High}	-0.008*** (0.002)
Inflation ² _{High}	0.018** (0.005)
Popgrowth ² _{High}	0.341** (0.104)
Government ² _{High}	0.003 (0.008)
_cons	11.027** (4.231)
N	1610
Justert R ²	0.1611

Standardfeil i parentes

* p < 0.05, ** p < 0.01, *** p < 0.001

Fra tabell 6 kan vi se hvilke parametere som er statistisk signifikante for lavinntektslandene.

Vi ser at sparing, befolkningsvekst og offentlig konsum forventes å være lineære i parameterne, mens inflasjon og Gini-koeffisienten forventes å ikke være lineære i parameterne. Fra tabell 7 ser vi at alle de uavhengige variablene utenom offentlig konsum er statistisk signifikante. Modellen for høyinntektsland bryter altså i stor grad med antakelsen om lineære parametere. De kvadrerte termene er med på å forklare den ikke-lineære sammenhengen mellom variablene, og ved å inkludere dem får vi en mer detaljert beskrivelse av den uavhengige variabelens effekt på den avhengige. Siden hovedfokuset i denne

oppgaven er hvordan Gini-koeffisienten påvirker BNP vekst per innbygger, velger vi kun å inkludere dens kvadrerte term i regresjonen. Funnet om de ikke-lineære sammenhengene er grunnlaget for at det kvadrerte begrepet ble inkludert i analysen i første omgang.

6.1.2 MLR.2 - Tilfeldig utvalg

Ifølge sentralgrenseteoremet vil et utvalg fra populasjonen på minst 30 etterleve kravet om et tilfeldig utvalg, noe vårt utvalg gjør (Anderson et al. 2020, s. 324). I vår modell har vi benyttet oss av tidsseriedata, noe som medfører komplikasjoner når det kommer til kravet om et tilfeldig utvalg. Et tilfeldig utvalg krever at hver observasjon, i , har den samme sannsynligheten for å bli valgt. Dette kravet brytes fordi tidsseriedata observerer de samme objektene over en gitt tidsperiode. Derfor kan vi ikke ha et tilfeldig utvalg av observasjoner, da det allerede finnes en naturlig temporal rekkefølge av observasjonene. Det er viktig å observere de samme objektene over tid ved bruken av tidsseriedata, fordi det inkluderer påvirkningen fortiden kan ha på fremtiden. For eksempel: årets Gini-koeffisient er påvirket av fjorårets Gini-koeffisient, som igjen ble påvirket Gini-koeffisienten året før der igjen og så videre.

Dette indikerer at MLR.2 ikke holder for vår modell. Likevel kan det gjøres justeringer som sørger for at MLR.2 holder ved bruken av tidsseriedata. Et eksempel er å sørge for stasjonaritet. Det krever at gjennomsnittet, variansen og kovariansen for enhver tidsforsinkelse er konstant over tid (Wooldridge, 2019, s. 345). For å konkludere er tidsseriedata problematisk med hensyn til MLR.2. Derfor bør konklusjoner fra denne studien tolkes med varsomhet. Forutsetning er likevel tatt hensyn til, og krenkelsen av MLR.2 blir problematisert videre i oppgavens del om robusthet og kritikk.

6.1.3 MLR.3 - Ingen perfekt kollinearitet

MLR.3 fordrer at de uavhengige variablene ikke kan være perfekt korrelerte, verken parvis eller gruppevis. I tilfeller med perfekt korrelasjon mellom variablene vil det være umulig å skille estimatene fra hverandre, fordi alle kombinasjoner av estimatene vil gi det samme resultatet (Wooldridge, 2019, s. 81). Vi bruker et standardisert mål for å avgjøre om variablene korrelerer i for stor grad. Dette målet sier at en høyere korrelasjon enn 0.9 (> 0.9) er uønsket (Wooldridge, 2019, 25 s. 92). I første omgang vil vi sjekke for kollinearitet ved bruk av en korrelasjonsmatrise. STATA brukes til å lage en korrelasjonsmatrise og undersøke

om det er noen verdier over terskelnivået. Tabell 8 viser korrelasjonsmatrisen for lavinntektsland.

Tabell 8: Korrelasjonsmatrise lavinntektsland

	<i>Gini</i>	<i>Gini</i> ²	<i>Savings</i>	<i>Inflation</i>	<i>Popgrowth</i>	<i>Government</i>
<i>Gini</i>	1.000					
<i>Gini</i> ²	0.989	1.000				
<i>Savings</i>	-0.219	-0.184	1.000			
<i>Inflation</i>	0.100	0.105	-0.061	1.000		
<i>Popgrowth</i>	.0280	0.305	0.467	-0.046	1.000	
<i>Government</i>	-0.558	-0.545	-0.305	-0.076	-0.173	1.000

Vi ser fra tabellen at *Gini* og *Gini*² er nesten perfekt korrelert, og vi har potensielt et problem. For å undersøke dette benytter vi oss av det statistiske målet variansens inflasjonsfaktor (VIF), som benyttes til å vurdere problemer knyttet til multikollinearitet (Ringdal og Viborg, 2017, s. 147). VIF-verdier på over 10 ansees som problematiske, det vil si at regresjonskoeffisientens standardfeil øker med en faktor på 10 (Ringdal og Viborg, 2017, s. 147). I den andre kolonnen vises den inverse av VIF. Toleransen defineres som $1 - R^2$ i en regresjon av X_i på de andre x-variablene. Vi ser at variablene offentlige utgifter, sparerate, befolkningsvekst og inflasjon har alle verdier over 1, men under 10. Dette indikerer at det finnes moderate korrelasjoner mellom en gitt forklaringsvariabel og andre forklaringsvariabler, men de korrelerer i så liten grad at vi ikke velger å ta stilling til dem. Tabellen viser derimot at både *Gini* og *Gini*² har VIF-verdier langt over kriteriet på 10. Likevel har vi allerede sett at både *Gini* og *Gini*² har statistisk signifikante regresjonskriterier, og estimatene av de andre koeffisientene i modellen påvirkes ikke i betydelig grad. Det vil heller ikke være et problem i vår modell da vi ikke vil se på effekten av å endre *Gini*-variabelen uten at vi endrer variabelen for *Gini*².

Tabell 9: VIF for lavinntektsland

Variabel	VIF	1/VIF
<i>Gini</i>	84.14	0.012
<i>Gini</i> ²	82.92	0.012
<i>Savings</i>	1.24	0.804
<i>Inflation</i>	1.01	0.987
<i>Popgrowth</i>	1.13	0.882
<i>Government</i>	1.27	0.787
Gj.nitt. av VIF	28.62	

Prosedyren er den samme for høyinntektsland Tabell 10 viser korrelasjonsmatrisen for høyinntektsland.

Tabell 10: Korrelasjonsmatrise høyinntektsland

	<i>Gini</i>	<i>Gini</i> ²	<i>Savings</i>	<i>Inflation</i>	<i>Popgrowth</i>	<i>Government</i>
<i>Gini</i>	1.000					
<i>Gini</i> ²	0.993	1.000				
<i>Savings</i>	-0.114	-0.111	1.000			
<i>Inflation</i>	-0.051	-0.043	-0.010	1.000		
<i>Popgrowth</i>	.0232	0.207	-0.127	0.036	1.000	
<i>Government</i>	-0.090	-0.080	0.429	-0.030	-0.030	1.000

Resultatene for høyinntektsland tilsvarer resultatene fra lavinntektsland og intuisjonen er stort den samme. Vi ser igjen at *Gini* og *Gini*² korrelerer. De samme variablene har også uønsket høye tall i VIF-testen, men vi velger likevel å inkludere dem fordi de har statistisk signifikante regresjonskriterier og korrelerer i liten grad med de andre variablene. Vi kan konkludere med at i store utvalg, slik som vårt, vil multikollinearitet som regel ikke være et problem.

Tabell 11: VIF for høyinntektsland

Variabel	VIF	1/VIF
<i>Gini</i>	60.74	0.016
<i>Gini</i> ²	58.76	0.017
<i>Savings</i>	1.25	0.800
<i>Inflation</i>	1.06	0.948
<i>Popgrowth</i>	1.20	0.835
<i>Government</i>	1.63	0.612
Gj.nitt. av VIF	20.77	

6.1.4 MLR.4 – Null-betinget gjennomsnitt

MLR.4 er vanskelig å oppfylle og direkte teste. Brudd på MLR.4 kommer som regel av utelatte variabler. Hvis det er en faktor i feilleddet som korrelerer med både den avhengige og den uavhengige variabelen, vil ikke forutsetningen om null-betinget gjennomsnitt være oppfylt, og estimatene vil ikke være objektive.

Det er grunn til å tro at vår modell bryter med denne forutsetningen. Modellen forsøker å predikere en sammenheng mellom inntektsulikhet og økonomisk vekst. Dette er en aggregert makroøkonomisk sammenheng som påvirkes av mange faktorer. Flere av disse er trolig ikke inkludert i vår modell. I oppgavens teoridel ble det vist til forskning som vektlegger imperfekte kredittmarkeder (Galor, 2012, s.26), samt den sosiopolitiske kanalen som viktige forklaringsfaktorer for økonomisk vekst (Alesina og Rodrik, 1994, s.1). Disse variablene ble utelatt fra regresjonsanalysen fordi det var vanskelig å finne god tilgjengelig data. Dette svekker modellens objektivitet. Forutsetningen har blitt etterlevd etter beste evne ved å inkludere så mange relevante kontrollvariabler som mulig.

6.1.5 MLR.5 - Homoskedasitet

MLR.5 omhandler feilbegrepets varians. For å finne ut om MLR.5 holder eller ikke vil vi benytte oss av en Breusch-Pagan-test (Wooldridge, 2019, s.270). Den gjennomføres i STATA, og tester om variansen i feilleddet er avhengig av verdiene i de uavhengige variablene. Testen kjøres både for utvalget av lavinntekts- og høyinntektsland.

Hypotesene er som følger:

H_0 : Homoskedasitet (de gjenværende er fordelt med lik varians)

H_A : Heteroskedasitet (de gjenværende er ikke fordelt med lik varians)

For utvalget av lavinntektsland er verdiene:

$$chi^2(1) = 19.88$$

$$prob > chi^2 = 0.00$$

For utvalget av høyinntektsland er verdiene:

$$chi^2(1) = 26.67$$

$$prob > chi^2 = 0.00$$

Testobservatorene er null for både utvalget av lavinntekt- og høyinntektsland. Med et signifikansnivå på 5% kan H_0 avises fordi p-verdiene er mindre enn 0,05. Vi kan konkludere med at variansen i feilbegrepene ikke er konstante og heteroskedasitet er til stede.

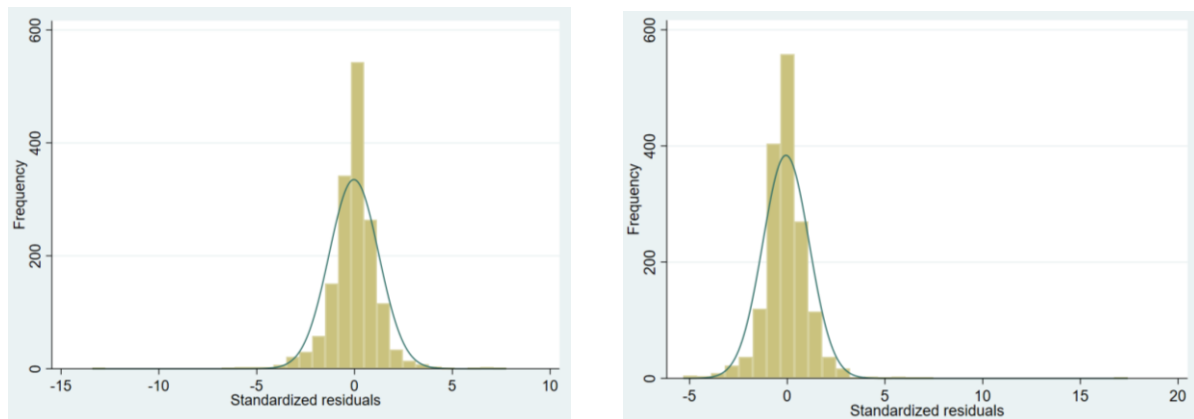
Regresjonen ser dermed ut til å bryte med MLR.5 og gir dermed upresise estimater av standardavvikene.

6.1.6 MLR.6 – Normalitetsantakelsen

MLR.6 forutsetningen er vanskelig å oppfylle, og etterleves sjeldent, da den fordrer at feilbegrepet er uavhengig av alle variablene, både ved forventet verdi og varians (Wooldridge, 2019, s. 118). På bakgrunn av sentralgrenseteoremet kan feilleddet sies å følge en normalfordeling. Dette kan gjøres på grunn av at feilleddet er summen av mange uobserverbare faktorer som påvirker den uavhengige variabelen (Wooldridge, 2019, s. 119). Disse faktorene kunne ha påvirket den avhengige variabelen. Selv om de ikke er inkludert, er det antatt at de vil – i tråd med sentralgrenseteoremet – påvirke den uavhengige variabelen separat og additivt (Wooldridge, 2019, s. 119-120).

Fordelingen av feilleddene er vist nedenfor. Figur 7 viser fordelingen for utvalget av lavinntektsland på venstresiden og utvalget av høyinntektsland på høyresiden. Linjen i figurene viser hvordan en normalfordeling ville sett ut, mens stolpene viser hvordan feilleddene faktisk er fordelt.

Figur 7: Normal fordeling for lav- og høyinntekts land



Grafisk ser det ut som fordelingen til både lavinntekts- og høyinntektsland er normalfordelt med null i gjennomsnitt. Intuitivt kan man konkludere i at MLR.6 holder. Likevel gjennomfører vi en statistisk test for å sjekke om feilleddenes avvik fra normalfordeling er statistisk signifikante. Det tas i bruk en sk-test som måler feilleddenes fordeling basert på mål for skjevhet og kurtosis, hvor skjevhet går på avvik fra symmetri i fordelingen og kurtosis forteller om fordelingen er flatere eller mer toppet enn normalfordelingen som har verdien 0 (Ringdal og Viborg, 2017, s. 75).

Hypotesene er som følger:

H_0 : Dataen følger en normalfordeling.

H_A : Dataen følger ikke en normalfordeling.

Utvalget for lavinntektsland:

$$chi^2(2) = 372.67$$

$$Prob > chi^2 = 0.00$$

Utvalget for høyinntektsland:

$$chi^2(2) = 764.17$$

$$Prob > chi^2 = 0.00$$

Med et signifikansnivå lik 5% og p-verdier lavere enn 0.05 avkreftes nullhypotesen til fordel for alternativhypotesen for begge utvalgene. De høye verdiene for kji kvadrert og de lave p-

verdiene tilsier at feilleddene ikke følger en normalfordeling. Avvikene fra normalfordelingen kan skyldes to forhold: selve avvikene og utvalgsstørrelsen. Funnene står i kontrast til hva vi tolket av den grafiske løsningen. Til forsvar for vår modell vil selv mindre avvik fra normalfordeling gi statistisk signifikante utfall (Ringdal og Viborg, 2017, s. 75).

7. Resultater og diskusjon

Hovedfunnene fra oppgavens analysedel er ulike for de to utvalgene. For lavinntektsland observerte vi en konkav signifikant sammenheng. For høyinntektslandene fant vi derimot en konveks signifikant sammenheng. Interaksjonsbegrepet i modell 3 underbygger begge disse funnene, men dette forholdet er ikke statistisk signifikant. Videre i diskusjonsdelen vil det trekkes det linjer mellom funnene fra regresjonsanalysen og de ulike teoretiske perspektivene. Dette vil diskuteres i lys av oppgavens problemstilling:

“Har inntektsulikhet en negativ eller positiv effekt på økonomisk vekst, og vil denne sammenhengen være forskjellig for høyinntekts- og lavinntektsland?”

Funnene for lavinntektsland kan sees i lys av Kuznets-kurven. Analysen vår viser en lignende kvadratisk sammenheng, men med en motsatt årsakssammenheng. Estimaten våre viser at inntektsulikhet er assosiert med økt økonomisk vekst, i avtakende grad, fram til et toppunkt med en estimert Gini-koeffisient lik 41.08. Etter dette er forholdet negativt.

Funnene for høyinntektsland kan på den andre siden sees i lys av Pikketys sammenheng. Han foreslår et konvekst forhold mellom faktorene. Resultatene våre viser et lignende konvekst forhold, hvor økt inntektsulikhet er assosiert med redusert vekst, i avtakende grad, fram til et bunnpunkt med en estimert Gini-koeffisient lik 43.38.

Det er viktig å merke seg at Pikketys og Kuznets årsakssammenheng går motsatt vei av det analysen vår estimerer. I likhet med dem ser vi på ikke-lineære sammenhenger mellom de to variablene, men funnene er ikke helt sammenlignbare. Inntektsulikhet kan dermed både anses som et resultat av økt økonomisk vekst og en faktor som påvirker økonomisk vekst. Begge teoriene viser likevel en korrelasjon med en styrke og en retning som deler flere likhetstrekk med funnene fra regresjonsanalysen vår.

Ifølge den klassiske tilnærmingen vil økt ulikhet være en forklarende faktor for økt vekst, og et resultat av høyere sparerate blant en rik befolkningsandel med en voksende inntektsandel

relativt til andre inntektsgrupper. Keynes og Kaldor peker mer spesifikt på kapitalakkumulering som følge av økt sparing som et viktig forhold for veksten i den post-industrielle perioden ((Keynes og Cox, 2019, s. 54), (Kaldor, 1957, s. 603)). Det kan muligens trekkes linjer mellom denne sammenhengen og den positive korrelasjonen vi observerer for lavinntektsland før toppunktet i den kvadratiske sammenhengen.

Den negative sammenhengen for lavinntektsland etter toppunktet kan sees i lys av teoriene om den sosio-politiske kanalen. Teoriene hevder at dersom spennet mellom rike og fattige blir for stort, vil det gi grobunn for politisk uro. Dette vil i neste omgang bidra til lavere økonomisk vekst. Videre foreligger det direkte kostnader knyttet til sosial misnøye og uro. Opprør og opptøyer resulterer til dårlig utnyttelse av ressursene i samfunnet. Det bidrar blant annet til et svekket insentiv for innovasjon og færre sysselsatte innbyggere, noe som bidrar til en bremsert vekst og utvikling i samfunnet. Forholdet støttes også opp under teorien om imperfekte kredittmarkeder, som belyser at fattige land med stor ulikhet vil ha flere individer som ekskluderes fra kredittmarkedene og dermed ikke får lån, noe som vil påvirke veksten negativt (Banerjee og Duflo, 2010, s. 62).

Teorien om et imperfekt kredittmarked kan også danne grunnlaget for forskjellene mellom de estimerte sammenhengene for lavinntekt- og høyinntektslandene. Galor hevder at fattige land er informelle utlånere, med relativt høye utlånsrenter. En konsekvens av dette er at utlånere ikke låner ut, selv om utlånet ville bidratt til økonomisk vekst (Galor, 2012, s. 26). I høyinntektsland er det derimot ofte sikrere informasjonsflyt og lettere gjennomførbare lånetransaksjoner, slik at folk kan ta opp lån for å investere i lønnsomme prosjekter. Dette illustrerer en sentral forskjell mellom lavinntekt- og høyinntektsland som kan være på å danne grunnlaget for de ulike sammenhengene vi har estimert i analysen.

8. Robusthet og kritikk

Denne analysen inneholder svakheter og begrensinger det må bli redegjort for. Regresjonsanalysen krevde at en rekke forutsetninger ble oppfylt for at den skulle være presis og objektiv. Flere av disse ble ikke oppfylt. Betydningen av brytningen med disse forutsetningene for modellen skal forklares. I det følgende skal vi gjennomgå regresjonens økonometriske svakheter og begrensninger, og kommentere analysens robusthet.

Utelatte variabler

De makroøkonomiske sammenhengene studien har sett på er store og innviklede. Økonomisk vekst er et komplekst mål som påvirkes av mye. Det finnes variabler som er med på å forklare variasjonen i økonomisk vekst på tvers av land som vi ikke har inkludert, som for eksempel kredittmarkeder og den sosiopolitiske kanalen.

Teorien om det imperfekte kredittmarkedet belyser at land med stor ulikhet vil ha flere individer som ekskluderes fra kredittmarkedene, siden de ikke får lån. Påvirkningen fra den sosio-politiske kanalen ble heller ikke inkludert. Det var krevende å finne gode datasett for kredittmarkeder og sosial uro. Verdensbankens innsamlede data var tynn, der det blant annet inneholdt store hull i dataen. Vi har derfor valgt å ikke ta med disse variablene i analysen.

Disse to og andre mulige variabler ble ikke inkludert i modellen. Utelatelsen skyldes blant annet problemer med å finne troverdig data over tidsperioden vi har valgt å se på. Vi har tidligere antatt at MLR.4 holder for å unngå en partisk modell, men ser på bakgrunn av de utelatte variablene av relevans kan gi en partisk modell (Wooldridge, 2019, s. 84-85).

Tilfeldig utvalg

Tidsseriedataen som benyttes bryter med MLR.4 om et tilfeldig utvalg. Den består av observasjoner som er satt opp i en kronologisk rekkefølge. Dette resulterer i at kravet om et tilfeldig populasjonsutvalg blir brutt. Dette er en svakhet i regresjonsanalysen – og generelt i bruken av tidsseriedata. Analysen kunne med fordel benyttet seg av utvalgsteknikker som tar stilling til dataens temporale struktur. Dette er noe leseren bør ta stilling til i sin tolkning av undersøkelsen og dens resultater.

Ingen seriell korrelasjon

Ettersom vi har brukt tidsserie-data er det viktig å ta tidsaspektet i betraktning. Vi må anerkjenne at tidligere observasjoner kan ha en effekt på de fremtidige observasjonene. En viktig antakelse for tidsserie-data som vi ikke har gått i dybden på er ingen seriell korrelasjon. Forutsetningen sier at feilleddene i forskjellige perioder ikke korrelerer. Dersom denne forutsetningen ikke holder vil estimatene i regresjonsanalysen vår gi unøyaktige konfidensintervaller (Wooldridge, 2019, s. 344 og 345). Dette betyr eksempelvis at uforventet høy inflasjon i en periode, sannsynlig gir nivåer over gjennomsnittet i neste periode. For å

finne ut hvorvidt vår analyse har seriell korrelasjon eller ikke, kunne vi brukt en Durbin-Watson test. Gitt at vi har så mange observasjoner og forskjellige variabler over en lang tidsperiode er det rimelig å anta at vi i flere tilfeller bryter med antagelsen.

Forsinkede variabler

Forsinkede variabler tar for seg den temporale korrelasjonen i de uavhengige variablene. I motsetning til feilleddene gir vi rom for temporal korrelasjon mellom variablene. Likevel er det viktig å ta stilling til viktigheten og begrensningene til forsinkede variabler. En forsinket variabel betyr at variabelens verdi er påvirket av tidligere perioder. Dagens offentlige konsum er med stor sannsynlighet påvirket av tidligere perioders offentlige konsum, og framtidens offentlige konsum er sannsynligvis et resultat av tidligere perioders konsum. For å oppsummere er neste års økonomiske vekst sannsynligvis et resultat av nåværende verdier for vekst og forsinkede uavhengige variabler. Verken forsinkede avhengige eller uavhengige variabler er inkludert i denne studien, noe som er en begrensning.

Null-betinget sannsynlighet/Problemet med simultan endogenitet

Allerede basert på litteraturen vi tar utgangspunkt var det en usikkerhet rundt hvilken retning det er på årsakssammenhengen mellom ulikhet og vekst. Det stiller spørsmål ved om det er økonomisk vekst som påvirker ulikheten eller om det er ulikheten som påvirker den økonomiske veksten. Dette slår tvil om interessevariabelen, inntektsulikhet, er eksogen, eller om den er endogent bestemt av deres avhengige variabel, vekst i BNP per innbygger. Dette kan skape empiriske utfordringer og slå tvil om resultatene deres.

Passform

Som forklart i 4.2 viser R^2 modellens passform. I utvalgene for lavinntektsland og høyninntektsland er den justerte R-kvadrert respektivt lik 0.15 og 0.131. Disse tallene er ikke tilfredsstillende, og et resultat av dette kan være at de uavhengige variablene ikke forklarer mye av variansen for BNP per innbygger. Dette er ikke nødvendigvis problematisk for modellens funn, men er likevel verdt å nevne.

9. Konklusjon

Denne bacheloroppgaven har undersøkt sammenhengen mellom inntektsulikhet og økonomisk vekst for høyinntektsland og lavinntektsland. Målet med oppgaven var å finne ut om graden av inntektsulikhet påvirker økonomisk vekst, og om det foreligger en ulik sammenheng for lav- og høyinntektsland.

Innledningsvis begynte vi med å presentere tidligere forskning og økonomisk teori. Den tidligere forskningen ga motstridende resultater. Videre anvender vi en multippel lineær regresjonsmodell for å analysere forholdet. Ved å dele inn utvalget i lav- og høyinntektsland, ble det gjennomført to ulike regresjoner. Regresjonen med lavinntektsland fant en konkav statistisk signifikant sammenheng mellom inntektsulikhet og økonomisk vekst, mens for høyinntektsland fant vi en konveks signifikant sammenheng. Avslutningsvis i analysedelen inkluderte vi et interaksjonsbegrep som viste en liknende ikke-signifikant sammenheng. I oppgavens diskusjonsdel ble resultatene diskutert i lys av teori og tidligere forskning.

Ut ifra våre resultater kan vi konkludere med at det foreligger ulik sammenheng mellom inntektsulikhet og økonomisk vekst for lavinntektsland og høyinntektsland. Derimot bør konklusjonene som trekkes fra analysen, tolkes med varsomhet. Modellen inneholder flere begrensninger som svekker resultatenes robusthet og objektivitet. Videre forskning bør ta stilling til bruken av tidsseriedata i større grad, slik at analysens funn kan tolkes med større gyldighet.

10. Litteraturliste

Alesina, A. og Rodrik, D. (1994). Distributive Politics and Economic Growth. *The Quarterly Journal of Economics*, 109(2), s. 465-490. Hentet fra <https://www.jstor.org/stable/2118470?seq=1>

Alesina, A. og Perotti, R (1996). Income distribution, political instability, and investment. *European Economic Review*, 40(6): 1203-1228. Hentet fra: https://dash.harvard.edu/bitstream/handle/1/4553018/alesina_incomedistribution.pdf

Alesina, A. og Rodrik, D. (1994). Distributive Politics and Economic Growth. *The Quarterly Journal of Economics*, 109(2), s. 465-490. Hentet fra <https://www.jstor.org/stable/2118470?seq=1>

Anderson, D. R., Camm, J. B., Cochran, J. J., Fry, M. J., Ohlmann, J. W., Sweeney, D. J., Williams, T. A. (2020). *Essential of Modern Business Statistics* (8. Utgave). Cengage Learning.

Banarjee, A. og Duflo, E. (2010). Giving credit where it is due. *The journal of Economic Perspectives*, 24(3):61-80. Hentet fra <https://www.jstor.org/stable/20799155>

Barro. R. (2003). Determinants of Economic Growth in a Panel of Countries. *Annals of Economics and Finance*, vol. 4(2), side. 231-274. Hentet fra: <http://down.aefweb.net/WorkingPapers/w505.pdf>

Bénabou, R. (1996). Inequality and Growth. *National Bureau of Economic Research*, Vol 11, s. 11 – 92. Hentet fra: <https://www.nber.org/system/files/chapters/c11027/c11027.pdf>

Canarella, G. & Pollard, S. K., 2003. The Augmented Solow Model and the OECD Sample. s.l.:International Business & Economics Journal.

Chirwa, T. G. og Odhiambo, N. M., 2016. Macroeconomic Determinants of Economic Growth: A review of International Literature. Vol 11(2) red. s.l.:De Gruyter.

Fengler, W. (2010, 15.april). *Can rapid population growth be good for economic development?* World Bank Blogs. Hentet fra: <https://blogs.worldbank.org/africacan/can-rapid-population-growth-be-good-foreconomic-development>

Galor, O. (2012). Inequality, Human Capital Formation and the Process of Development. Brown University and IZA, No. 6328. Hentet fra: <https://docs.iza.org/dp6328.pdf>

Kaldor, N. (1957). A Modell of Economic Growth. *Economic Journal*, 67(268), s. 591-624. Hentet fra <https://www.jstor.org/stable/2227704>

Keynes, J. M. og Cox, M. (2019). *The Economic Consequences of the Peace: With a new introduction by Michael Cox*. Cham: Springer International Publishing.

Kuznets, S. (1955). Economic Growth and Income Inequality. *The American Economic Review*, 45(1), s. 1-28. Hentet fra <http://www.jstor.org/stable/1811581>

Meltzer, A. og Richards, S. (1981). A rational theory of the size of government. *Journal of Political Economy*, 89(5), s. 914-927. Hentet fra <https://www.jstor.org/stable/1830813?seq=1>

NHO (2018). Verden og oss: Næringslivets perspektivmelding 2018. Hentet fra: https://www.nho.no/siteassets/publikasjoner/naringslivets-perspektivmelding/pdf-er-okt18/nho_perspektivmeldingen_6_okende-ulikhet.pdf

Norges Bank. (2019, 10.april). Hvorfor vil vi ha lav og stabil inflasjon? Hentet 20. mars fra <https://www.norges-bank.no/kunnskapsbanken/inflasjon/hvorfor-vil-vi-ha-lav-og-stabil-inflasjon/>

Persson, T. og Tabellini, G. (1994). Is Inequality Harmful for Growth? *The American Economic Review*, 84(3), s. 600-621. Hentet fra <https://www.jstor.org/stable/2118070?seq=19>

Ram, R. (1986). Government Size and Economic Growth: A New Framework and Some Evidence from Cross-Section and Time-Series Data. *The American Economic Review*, 76(1), s. 191- 203. Hentet fra: <https://www.jstor.org/stable/1804136>

Ringdal, K og Wiborg, Ø. (2017). *Lær deg Stata*. (1. Utgave). Fagbokforlaget.

Roser, M. (2013). Economic Growth. *Our World In Data*. Hentet fra:

<https://ourworldindata.org/economic-growth>

The World Bank. *Domestic credit provided by financial sector (% of GDP)*. Hentet 21.mars 2023 fra <https://data.worldbank.org/indicator/.AST.DOMS.GD.ZS.AST.DOMS.GD.ZS>

The World Bank (2023). "GNI per capita, PPP (current international \$)".

<https://data.worldbank.org/indicator/NY.GNP.PCAP.PP.CD>

The World Bank (2023). "*World Bank Country and Lending Groups*". Hentet 21.mars fra

<https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups>

Todaro, M. og Smith, S. (2020) *Economic Development*, (13. Utgave), Pearson Ltd.

<https://data.worldbank.org/about>

Wooldridge, J.M. (2015). *Introductory Econometrics A Modern Approach* (6. Utgave).

Cengage Learning.

