

Kristian Fliflet Sollie, Edvard Kilvik Hoff, Julian Pedersen Grøen, Marcus Solstrand Fossli

Analyse av arbeidsledighetens effekt på vinningskriminalitet

Bacheloroppgave i Samfunnsøkonomi

Veileder: Kåre Johansen

Mai 2023

Kristian Fliflet Sollie, Edvard Kilvik Hoff, Julian
Pedersen Grøen, Marcus Solstrand Fossli

Analyse av arbeidsledighetens effekt på vinningskriminalitet

Bacheloroppgave i Samfunnsøkonomi
Veileder: Kåre Johansen
Mai 2023

Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
Institutt for samfunnsøkonomi



Kunnskap for en bedre verden

Forord

Denne bacheloroppgaven markerer slutten på en viktig fase i vårt utdanningsløp.

Takk til Kåre Johansen, vår veileder, for hans veiledning og støtte.

Edvard Kilvik Hoff, Kristian Fliflet Sollie, Julian Pedersen Grøen og Marcus Solstrand Fossli.

15. mai 2023

Sammendrag

Formålet med denne oppgaven er å analysere påvirkningen arbeidsledighet har på omfanget av vinningskriminalitet i Norge. Oppgaven skal også ta for seg andre kontrollvariabler som er forventet å ha effekt og forsterke analysen. Vi gjør nytte av en OLS-regresjonsanalyse for å undersøke relasjonene mellom variablene. Gjennom analysen ser vi på den direkte effekten av arbeidsledighet på vinningskriminalitet og dens effekt korrigert for vedvarende årlig nedgang i vinningskriminalitet. Vi bruker to ulike økonometriske modeller i denne oppgaven; en hovedmodell og en supplerende logaritmisk modell. I regresjon av begge modellene har vi funnet en statistisk signifikant og positiv sammenheng mellom arbeidsløshet og vinningskriminalitet. Vi har også funnet signifikante relasjoner mellom vinningskriminalitet og flere av våre kontrollvariabler.

Abstract

The purpose of this thesis is to analyze the impact of unemployment on the extent of property crime in Norway. The assignment will also deal with other control variables that are expected to have an effect and strengthen our analysis. We make use of an OLS regression analysis to examine the relationships between the variables. Through the analysis, we look at the direct effect of unemployment on commercial crime and its effect corrected for a sustained annual decline in commercial crime. We use various econometric models in this assignment; a main model and a supplementary logarithmic model. In the regression of both models, we have found a statistically significant and positive correlation between unemployment and property. We have also found significant relationships between property crime and several of our control variables.

Contents

1	Innledning og problemstilling	5
1.1	Bakgrunn	5
1.2	Problemstilling	5
2	Litteraturgjennomgang	6
2.1	Teorier knyttet til kriminalitet	6
2.2	Resultater fra tidligere studier	7
3	Datagrunnlag	8
3.1	Datainnsamling	8
3.2	Variabler	8
3.2.1	Total vinningskriminalitet	8
3.2.2	Arbeidsledighet	9
3.2.3	Langtidsledighet	10
3.2.4	Prosentvis endring i BNP	11
3.2.5	Lavinntekt	12
3.2.6	Prosentandelen av innvandrere i befolkningen	13
3.2.7	Andel oppklarte eiendomstyveri	14
3.3	Korrelasjon og korrelasjonsmatrise	15
4	Metode	16
4.1	Regresjon	16
4.2	Statistikkmetode: Minste kvadraters metode	16
4.3	Goodness of fit	18
4.4	Variansinflasjonsfaktor (VIF)	18
4.5	Hypotesetesting og signifikans	19
4.6	Modellene	21
5	Resultater	22
5.1	Enkel lineær regresjon av vinningskriminalitet og arbeidsledighet	22
5.2	Resultater fra hovedmodellen	22
5.2.1	Trend (År)	23
5.2.2	Arbeidsledighet	23
5.2.3	Langtidsledighet	23
5.2.4	Endring i BNP	24
5.2.5	Oppklaringsprosent av eiendomstyveri	24
5.2.6	Prosent innvandrere i befolkningen	24
5.2.7	Lavinntekt	24
5.3	Tolkning av resultater	24
5.4	Enkel lineær regresjon av vinningskriminalitet og år	25
5.5	Enkel lineær regresjon med trendjustert vinningskriminalitet	25
5.6	Resultater fra alternativ modell	26
5.6.1	Arbeidsledighet	27
5.6.2	Langtidsledighet	27
5.6.3	Endringen i bruttonasjonalprodukt	27
5.6.4	Prosentandelen av innvandrere i befolkningen	27

5.6.5	Lavinntekt	27
5.6.6	Oppklaringsprosent	28
5.7	Hypotesetesting av variabler i modellene	28
5.8	VIF	29
5.9	Sammenligning av resultater fra hovedmodell og trendjustert logaritmisk modell	30
5.10	Kritikk og robusthet	30
6	Konklusjon	31
7	Referanseliste med kilder og statistikkbank	32
7.1	Kilder	32
7.2	Statistikkbank	33
8	Appendiks	34

1 Innledning og problemstilling

”Vinningslovbrudd er lovbrudd som begås i den hensikt å skaffe seg eller andre goder som penger eller verdifulle gjenstander” (Lomell, 2021)

1.1 Bakgrunn

Arbeidsledighet vil til enhver tid være tilstedeværende i økonomien. Kontinuerlig omstilling og teknologisk utvikling er en av grunnene til dette. Myndighetene har ikke mål om en ledighetsprosent på null, men fører en aktiv politikk for å holde ledigheten på et lavt og stabilt nivå. Vedvarende høy arbeidsledighet er derimot dårlig for økonomisk vekst og økonomisk velstand. Aggregert produksjon havner på et ugunstig lavt nivå, og konsumet faller fordi mange mister jobben, og får mindre å rutte med. I tillegg vil det tære på statskassen, fordi de ledige trenger kompensasjon for inntektstapet. De makroøkonomiske konsekvensene av uønsket nivå på arbeidsledigheten er altså tydelige. De sosiale konsekvensene, i vårt tilfelle økt vinningskriminalitet, er trolig ikke like åpenbare, og kan ikke nødvendigvis forklares utelukkende av standard økonomisk teori.

Rent intuitivt kan man tenke seg flere grunner til hvorfor økt ledighet i befolkningen fører til økt kriminalitet. Har man ikke fast arbeid å gå til, er det godt mulig at man fortere kjeder seg, noe som igjen kan føre til at fristelsen til å bryte loven blir større. Dessuten er det ikke utenkelig at lediggang fører til utenforskap og en følelse av å ikke være like deltakende i storsamfunnet som sysselsatte. Da blir muligens terskelen for å begå kriminelle handlinger lavere. Når det gjelder vinningskriminalitet spesifikt, kan tap av lønnsarbeid være en motivator for å begå innbrudd og lignende, slik at man sikrer seg en inntekt. Vår oppgave tar utgangspunkt i det norske samfunnet. Mister man jobben i Norge, er det myndighetenes ansvar å sørge for at man fortsatt har en inntekt ved hjelp av offentlige ytelser (Arbeids- og inkluderingsdepartementet, 2023). Dermed vil kanskje ikke ledighet ha en like stor effekt på vinningskriminalitet, sammenlignet med andre land hvor tap av arbeidsinntekt kan gi større økonomiske konsekvenser for individer og husholdninger.

1.2 Problemstilling

I samsvar med den underliggende teorien og eksisterende litteratur om sammenhengen mellom vinningskriminalitet og arbeidsledighet, vil dette prosjektet undersøke forholdet mellom vinningskriminalitet og arbeidsledighet gjennom en 19-års periode fra 2003 til 2021. For å oppnå dette, benyttes tidsseriedata på makronivå med årlig data frekvens for Norge for de aktuelle årene, og økonomiske metoder anvendes for analyse. Problemstillingen i denne oppgaven er:

Hvilken påvirkning har arbeidsledighet på vinningskriminalitet?

2 Litteraturgjennomgang

I dette kapitlet vil vi presentere noen relevante teorier, samt gjennomgå resultater fra et utvalg tidligere studier som har forsøkt å svare på problemstillinger lik vår egen.

2.1 Teorier knyttet til kriminalitet

Forholdet mellom kriminalitet og arbeidsledighet har lenge vært et forskningstema. Raphael & Winter-Ebmer (2001) analyserer sammenhengen mellom arbeidsledighet og kriminalitet i det amerikanske samfunnet. De bruker data fra samtlige delstater med unntak av Washington D.C. I det teoretiske rammeverket for studien tar forskerne utgangspunkt i tanken om at det å begå kriminelle handlinger som gir økonomiske gevinster kan bli betraktet som en type sysselsetting på lik linje med lovlige arbeid. Basert på en allerede etablert modell av den amerikanske professoren Jeffrey Grogger, betrakter de kriminalitet på et individnivå/mikronivå, hvor hver aktør tar avgjørelser basert på hvor mye nytte de får av henholdsvis ulovlige og lovlige sysler i forhold til hvor mye tid de må vie. Modellen hevder at potensielle lovbrøyttere vil unngå kriminelle handlinger med lav avkastning først, og heller prioritere de som gir stor gevinst. Dermed vil nytten av vinningskriminalitet avta etterhvert som man begår flere og flere lovbrudd. I samme studie mener de at lønnsnivået i forhold til den marginale nytten av vinningskriminalitet vil påvirke hvordan ulike individer bruker tiden sin på å generere inntekt. Hvordan et individ reagerer på å bli arbeidsledig avhenger av hvordan individet bruker tiden sin i utgangspunktet. Noen lønnsstakere har såpass høye arbeidsinntekter at marginalnyttens av kriminell aktivitet aldri vil være større enn marginalnyttens av arbeid. De vil av den grunn ikke sette av en viss mengde tid til å utføre ulovlige aktiviteter med økonomiske insentiver. Hvis de imidlertid mister jobben, kan dette endre seg. Her trekker Raphael og Winter-Ebmer inn begrepet reservasjonslønn. Dette er det minimale lønnsnivået en arbeider kan bli tilbudt for at vedkommende skal være totalt likegyldig mellom å velge å ta en jobb eller ikke ta den. De med forholdsvis lave reservasjonslønner, vil ifølge modellen være mer tilbøyelig til å begå kriminalitet om de mister arbeidsinntekt. Det motsatte gjelder for de med høye reservasjonslønner. Skulle det være slik at man ikke har jobb og bare begår kriminelle handlinger, vil man ikke uansett bli rammet av arbeidsledighet. Derfor vil man heller ikke bidra til at omfanget av vinningskriminalitet øker på makronivå hvis den generelle arbeidsledigheten tiltar. For individer som skaffer seg inntekter fra kriminalitet og inntekter fra normalt lønnsarbeid, vil tap av arbeid sannsynligvis føre til at man begår mer kriminalitet enn tidligere (Raphael & Winter-Ebmer, 2001, s. 262-264).

Modellen Raphael og Winter-Ebmer har som utgangspunkt for sin empiriske forskning, er en typisk økonomisk modell med antatt rasjonelle aktører som står overfor ulike valg av hvordan enhver person disponerer sin tid. Det kan være relevant å knytte økonomisk teori til kriminalitet, siden kriminalitet ofte blir begått med hensikt om å øke sin økonomiske nytte. Samtidig kan det også være lurt å trekke inn teorier som ikke nødvendigvis hviler på antagelsen om at kriminelle handler rasjonelt, og som i større grad vektlegger psykologiske faktorer. Et godt eksempel er general strain theory, som er utviklet av kriminologen Dr. Robert Agnew. Teorien hevder at en økning av stressende forhold og strains (norsk: press) hos et individ kan føre til at dette individet velger å begå kriminalitet for å få utløp for eller hanskes med stresset. Dette skjer blant annet fordi press ofte reduserer selvfølelse og psykisk velvære eller at press og negative sosiale relasjoner eller omgivelser senker terskelen for å begå kriminalitet og bryte normer. Agnew mener dette gjelder for generell kriminalitet, ikke bare vinningskriminalitet. Likevel understreker han at arbeidsledighet, spesielt langvarig arbeidsledighet, kan føre til færre lovlige framgangsmåter for å skaffe seg goder på. Dermed blir ikke eiendomstyveri og lignende lovbrudd fordømt i like stor grad hos individet. Det kan i tillegg kompensere for negative emosjoner og en følelse av å ikke strekke til. (Agnew & Brezina, 2019, s. 150-165). Hvis GST samsvarer godt med virkeligheten, kan vi tenke oss at det er

sannsynlig at langtidsledighet påvirker kriminalitet. Det er dog ikke sikkert hvor stor andel av de som eventuelt begår kriminalitet som følge av dårlige jobbmuligheter, faktisk er registrert som arbeidsledige. Det kan godt være at mange av disse individene for lengst har gitt opp på å lete etter muligheter for arbeid.

2.2 Resultater fra tidligere studier

I 2005 brukte den svenske økonomen Karin Edmark paneldata fra sitt eget hjemland i perioden 1988 til 1999 for å nærmere undersøke hvordan omfanget av “property crime” (som her omfatter ran, innbrudd, sykkel- og biltyveri, svindel og tyveri fra butikker og kjøretøy) ble påvirket av arbeidsledighetsraten. Andre variabler utover ledighet Edmark inkluderte var blant annet oppklaringsrate, gjennomsnittsinntekt og andelen utenlandske innbyggere. Edmark konkluderte med at ledighet hadde en tydelig positiv effekt på biltyveri og innbrudd, som var statistisk signifikant på både 10 og 5 prosent signifikansnivå. Dette var gjeldende for alle spesifikasjoner av modellen, også når det ble justert for tidstrender i de ulike svenske fylkene som ble undersøkt. Når det gjaldt andre typer property crime, fant man ingen betydelig sammenheng med arbeidsledighet, med unntak av sykkeltyveri i enkelte modell-spesifikasjoner. Dette var også tilfelle når man så på voldelige kriminelle handlinger, som var en annen hovedkategori (Edmark, 2005, s. 353-373). Det skal imidlertid sies at data Edmark bruker i sin analyse er hentet fra et tidsintervall som ligger utenfor vårt eget, og tar for seg et annet land. På den annen side kan det faktum at det er relativt små sosioøkonomiske og kulturelle forskjeller mellom Sverige og Norge tale for at Edmark sine funn er interessante for vår oppgave, siden de kan indikere lignende forhold i det norske samfunnet.

Ved hjelp av OLS-metoden estimerte professor Duha T. Altindag i 2012 at en økning i arbeidsledighetsraten med ett prosentpoeng ville føre til en forventet økning av “total property crime” med 1,7%. Her er property crime definert som tyveri, innbrudd og stjeling av kjøretøy. Relasjonen er signifikant på et 1% nivå. Dette betyr at sannsynligheten for at disse resultatene ikke er tilfeldige er 99% eller høyere. Data-materialet til Altindag består av kriminalstatistikk fra 1996 til 2003 fra 33 europeiske land, skjønt hvilke år eller perioder som er dekket varierer fra land til land. Altindag har brukt mange kontrollvariabler; BNP per kapita og alkoholkonsum er to eksempler (Altindag, 2012, s. 145-157).

I den tidligere nevnte studien til Winter-Ebmer & Raphael (2001) fant de forøvrig et tydelig positivt forhold mellom vinningskriminalitet og arbeidsledighet, men ikke mellom arbeidsledighet og voldelige lovbrudd (Raphael & Winter-Ebmer, 2001, s. 22-23).

De ulike studiene ovenfor har ulike datagrunnlag. Likevel peker de respektive funnene alle i samme retning, nemlig at arbeidsledighet har en klar signifikant og positiv effekt på vinningskriminalitet.

3 Datagrunnlag

Dette kapitlet vil gjøre rede for hvordan vi har samlet inn data og hvilken statistikk vi har hentet for å lage både avhengig og uavhengige variabler i vår modell. I tillegg vil vi forklare enkelte funn i datagrunnlaget som er relevante for utforming og tilpassing av de økonometriske modellene vi bruker, samt definere og beskrive noen viktige termer og fenomener knyttet til vår oppgave.

3.1 Datainnsamling

I datasettet som benyttes i vår statistiske analyse er hentet utelukkende fra statistikkbanken til Statistisk sentralbyrå. Å benytte data som er samlet fra SSB sin statistikkbank gir oss visse fordeler. Statistisk sentralbyrå er et offentlig organ underlagt Finansdepartementet. Organet er en faglig uavhengig institusjon ansvarlig for å samle inn, produsere og publisere offisiell statistikk relatert til økonomi, befolkning og samfunn på nasjonalt, regionalt og lokalt nivå. SSB bedriver i tillegg omfattende forsknings- og analysevirksomhet. Faglig uavhengighet innebærer at organet selv bestemmer hva som publiseres. Organets oppdrag er viktig for et velfungerende demokrati og er avgjørende for debatt og forskning. Kravene til offisiell statistikk er fastsatt i statistikkloven og følger standardiserte prosessmodeller (FNs Generic Statistical Business Process Model (GSBPM)) som er nødvendige for å produsere offisiell statistikk (SSB, 2023). Dette sikrer at dataen har høy kvalitet og høy grad av reliabilitet.

3.2 Variabler

Variablene vi har valgt er total anmeldt vinningskriminalitet som vår avhengige variabel, eller y-variabel. Videre har vi arbeidsledighetsraten som vår uavhengige variabel og følgende kontrollvariabler: År, langtidsledighet, endring i BNP, andel med lav inntekt, prosent innvandrere i befolkningen og oppklaringsraten til Politiet i anmeldte eiendomstyverier.

3.2.1 Total vinningskriminalitet

Total vinningskriminalitet er en kvantitativ, kontinuerlig avhengig variabel som kvantifiserer det samlede antallet vinningskriminelle lovbrudd anmeldt per år. Denne variabelen er relevant for å utforske forholdet mellom økonomiske forhold og kriminalitet i samfunnet, hvor vinningskriminalitet fungerer som den avhengige variabelen som kan bli påvirket av ulike faktorer. Total vinningskriminalitet omfatter et utvalg av lovbrudd som er definert i straffeloven, for eksempel tyveri, underslag, ran, utpressing, heleri, hvitvasking og uautorisert bruk av kjøretøy (lovdata). Fellesnevneren for disse forbrytelsene er at de er drevet av økonomisk gevinst. Tidsserien for total vinningskriminalitet strekker seg fra 2003 til 2021 [1], med årlige intervaller, noe som gir et omfattende datagrunnlag for å analysere trender og mønstre i vinningskriminalitet over tid. Dataene på vinningskriminalitet som er hentet ut fra SSB omfatter følgende:

Eiendomstyveri: Tyveri i alt og annen bruk og besittelse av andres eiendel (SSB, 2017).

Annet vinningslovbrudd: Bedrageri, underslag, økonomisk utroskap, korrupsjon og påvirkningshandel, regnskapsovertredelse, kreditorrelatert lovbrudd, skatt og avgift, tolløven, annet næringslivs- og økonomilovbrudd (SSB, 2017).

En begrensning med denne avhengige variabelen er at den bare måler anmeldte lovbrudd og ikke nødvendigvis de faktiske lovbruddene. Dette innebærer at total vinningskriminalitet kan være underrapportert, ettersom det er mulig at en betydelig andel av kriminaliteten ikke blir anmeldt. Dette kan skyldes ulike faktorer, som frykt for represalier, mangel på tillit til politiet, eller at offeret ikke er klar

over at et lovbrudd har funnet sted. Det er også risiko for overrapportering, for eksempel i tilfeller av forsikringssvindel. Når variabelen er behandlet i datasett og STATA har den navnet *Vinningskriminalitet*.

Det har vært en klar negativ trend i både vinningskriminalitet og kriminalitet i sin helhet de siste tiårene. Vi blir nødt til å justere modellene for denne trenden i oppgaven vår ettersom den ikke nødvendigvis stammer fra relevante variabler som vi skal undersøke, samt at den vil gi negative tall i koeffisientene for alle variablene dersom det ikke korrigeres for. Hvordan vi gjør dette blir dekket senere i oppgaven. Under ser vi den årlige nedgangen i kriminalitet fra 2003 til 2021. Som vi kan se er det en klar negativ trend med noen små hopp i 2009 og 2013.

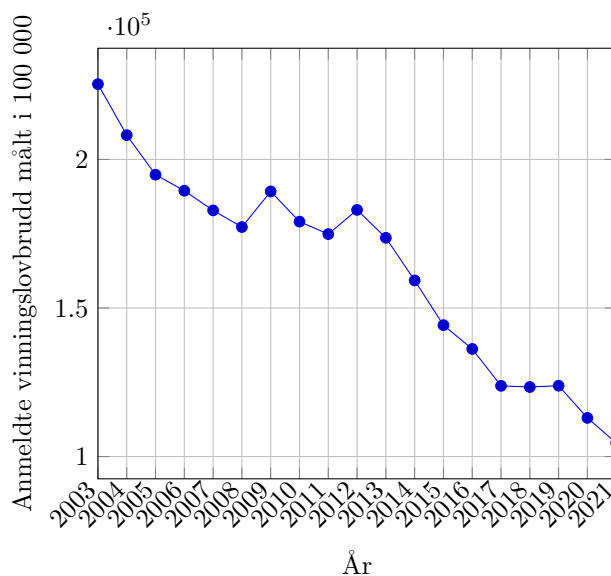


Figure 1: Vinningskriminalitet (2003-2021)

Trenden er godt dokumentert i mange land og kan trolig forklares blant annet av en underliggende årsak rundt lavere markedspriser for forbruksvarer på tvers av land, samt økte investeringer i sikkerhetstiltak. Etterhvert som samfunnet har blitt stadig mer digitalisert har også flere vinningskriminelle sett seg til å gå over til den digitale plattformen og prøve å skaffe seg ulovlige inntekter på denne måten. “Det er grunn til å anta at anmeldelsestilbøyeligheten for digital vinningskriminalitet er lav og henger antakelig sammen med størrelsen på økonomiske tap og oppfatninger om politiets etterforskningsevne” (Meld. St. 29 (2019-2020), s. 10-11).

3.2.2 Arbeidsledighet

Arbeidsledighet er en viktig økonomisk indikator som viser andelen av arbeidsstyrken som ikke har arbeid. Dataene vi benytter oss av er hentet fra arbeidskraftundersøkelsen til Statistisk sentralbyrå og gjelder for årene 2003 til 2021 [2]. Dette er tidsseriedata for arbeidsledige, som er andelen av arbeidsstyrken uten arbeid når undersøkelsen ble gjennomført. For å unngå effekten av faste sesongvariasjoner i sysselsetting, har vi valgt å bruke sesongjusterte data. Dette betyr at eventuelle sesongvariasjoner, som for eksempel utløp av kontrakter over nyttår, er justert for slik at tallene bedre representerer den faktiske arbeidsledigheten på årsbasis. Interpolering gjør det mulig å benytte mer detaljerte data og dermed få et mer nøyaktig og fullstendig bilde av arbeidsledigheten på årsbasis. Ved å bruke interpolering kan vi unngå å ta utgangspunkt i kun noen få datapunkter, for eksempel tallene for første kvartal, og dermed få en mer helhetlig forståelse av utviklingen i arbeidsmarkedet over tid. Dette gir oss muligheten til å trekke mer pålitelige og presise konklusjoner om arbeidsledigheten og hvordan den påvirkes av ulike faktorer.

Definisjonen av arbeidsledighet vi opererer med her er “personer uten inntektsgivende arbeid som aktivt forsøkte å skaffe seg arbeid i løpet av de siste fire ukene, og som kunne ha påtatt seg arbeid i løpet av referanseuka eller de to påfølgende ukene. Personer som har vært eller forventer å være permittert i mer enn tre måneder regnes også som arbeidsledige dersom de oppfyller disse kriteriene.” (SSB, 2023). Arbeidsledighet regnes i prosent av arbeidsstyrken som i arbeidskraftundersøkelsen er alle i alderen 15-74 år som er sysselsatt eller arbeidsledige. Statistikken gir også informasjon om de som står utenfor arbeidslivet. Merk at personer utenfor arbeidsstyrken regnes ikke som arbeidsledige. For å regnes som en del av arbeidsstyrken må en aktivt tilby arbeidskraften sin på arbeidsmarkedet. Personer uten arbeid som ikke har søkt jobb i løpet av de siste fire ukene da undersøkelsen ble gjennomført vil ikke regnes som arbeidsledig. Andre grupper som ikke regnes som arbeidsledige er personer under utdanning, pensjonerte arbeidsuføre og hjemmearbeidende. (SSB, 2023). Under ser vi utviklingen i arbeidsledighet i prosent. Arbeidsledigheten fluktuerer rundt et gjennomsnitt på ca. 3.97% med et standardavvik på 0.59. Toppunktet finner vi i 2017 på 4.9% arbeidsledige. I datasett og regresjonsanalyse gjort i STATA er variabelen behandlet med variabelnavnet *arblid*.

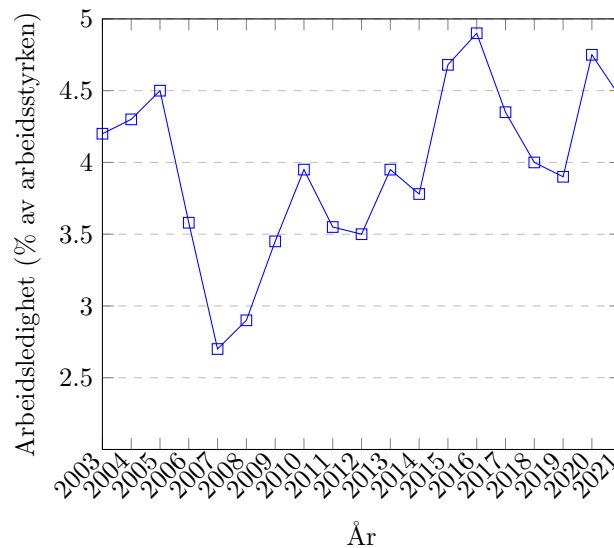


Figure 2: Arbeidsledighet (2003-2021)

3.2.3 Langtidsledighet

Å fokusere på langtidsledighet betyr å vurdere de som har vært uten arbeid over en betydelig tidsperiode, typisk målt i uker eller måneder. I Norge er det Arbeids- og velferdsdirektoratet (NAV) som definerer langtidsledighet som en tilstand der en person har vært arbeidsledig i 26 uker eller mer. For å analysere langtidsledigheten, har vi valgt å benytte data fra Statistisk sentralbyrås (SSB) arbeidskraftundersøkelse. SSB kategoriserer langtidsledighet som individer uten arbeid i 7-12 måneder, eller over et år. For en grundigere analyse, kombinerer vi SSBs langtidsledighet inkludert personer uten arbeid i 7-9 måneder, 10-12 måneder og lengre enn ett år [3]. Gitt at vår avhengige variabel er årlig, har vi valgt å interpolere tallene for langtidsledighet, slik at vi kan inkludere dem i analysen. Denne metoden lar oss estimere verdiene for de umålte månedene, og gir oss et mer helhetlig bilde av langtidsledigheten på årsbasis. På den måten kan vi gjennomføre mer detaljerte analyser av langtidsledighetens utvikling over tid og undersøke hvilke faktorer som påvirker denne. Under ser vi utviklingen i langtidsledigheten fra 2003 til 2021. Gjennomsnittlig utvikling over disse 19 årene er ca. 30% av arbeidsledige, standardavvik er ca. 5.49 med toppunkt på 37.5% i 2017.

I datasett og regresjonsanalyse gjort i STATA er variabelen behandlet med variabelnavnet *langed*.

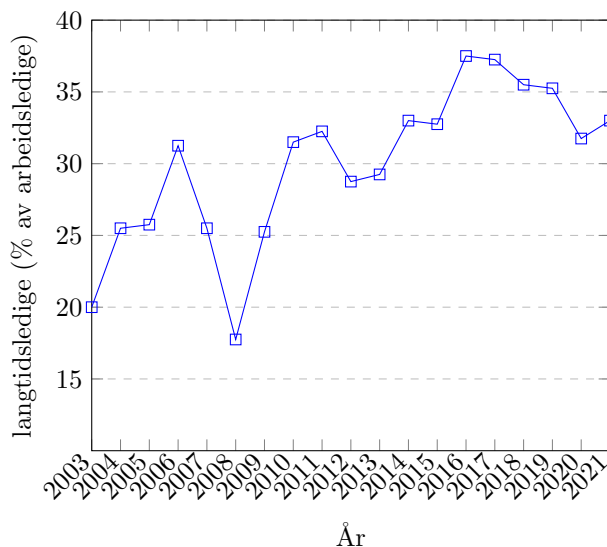


Figure 3: Langtidsledige, prosent av arbeidsledige (2003-2021)

3.2.4 Prosentvis endring i BNP

Endring i brutto nasjonalprodukt (BNP) er en viktig økonomisk indikator som måler veksten i et lands økonomi. Denne variabelen representerer prosentvis endring i BNP over tid, og gir innsikt i den generelle økonomiske ytelsen og utviklingen i et land. Tallene er gitt kvartalsvis, og vi bruker data fra 2003 til 2021 [4]. Inkluderingen av prosentvis endring i BNP i analysen har som formål å fange opp variasjonen i kriminaliteten som kan være relatert til økonomiske forhold. Ved å inkludere prosentvis endring i BNP som en uavhengig variabel, kan man undersøke hvordan endringer i økonomisk vekst påvirker kriminalitetsnivået, og dermed få en bedre forståelse av sammenhengen mellom økonomiske forhold og kriminalitet i samfunnet. Det er imidlertid viktig å merke seg at prosentvis endring i BNP kun gir en aggregert måling av økonomisk vekst, og kan derfor overse lokale eller sektorspesifikke variasjoner i økonomisk aktivitet. For å få et mer detaljert bilde av hvordan økonomiske forhold påvirker kriminalitet, kan det være nyttig å supplere denne variabelen med andre økonomiske indikatorer, som arbeidsledighetsrater, inntektsulikhet og sosiale støtteordninger. Sekundærhypotese: Økt vekst i BNP fører til mindre vinningskriminalitet. Under ser vi endringen i brutto nasjonalprodukt i prosent. Vi ser at det her er en svak positiv trend med unntak av noen daler rundt finanskrisen i 2008-2009, oljeprisfallet i 2014 og koronaepidemien i 2020. Veksten i BNP ligger på et gjennomsnitt på ca. 1.5% og et standardavvik på ca. 1.5. I datasett og regresjonsanalyse gjennomført i STATA er variabelen behandlet med variabelnavnet *BNP*.

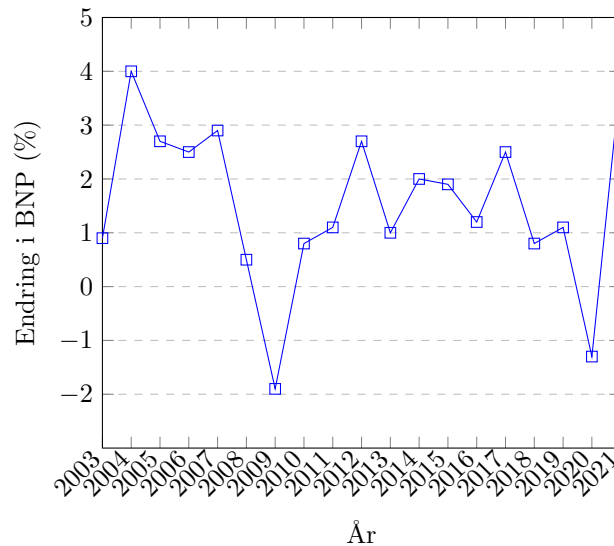


Figure 4: Ending i BNP (2003-2021)

3.2.5 Lavinntekt

Det er ingen fastsatt fattigdomsgrense i Norge. Derfor har vi tatt utgangspunkt i en lavinntektsgrense som Statistisk sentralbyrå (SSB 2023) bruker for å måle inntektsnivået til personer og husholdninger som kan anses som økonomisk svake. Grensen for lavinntekt, også kjent som "lavinntektsgrensen", settes på 60% av medianinntekten i Norge etter skatt, over en treårsperiode. Medianinntekten representerer den midterste verdien i et tallmateriale, det vil si lønnen til en person som befinner seg nøyaktig i midten av inntektsfordelingen i Norge [5]. Lavinntektsgrensen, basert på 60% av medianinntekten, gir et relativt mål på fattigdom som tar hensyn til inntektsnivået og levestandarden i det norske samfunnet. Dette relativt målet gjør det mulig å sammenligne inntektsforhold over tid og mellom ulike befolkningsgrupper, og det kan bidra til å identifisere de mest sårbare individene og familiene som kan trenge sosial støtte og hjelp. Ved å inkludere lavinntektsgrensen i en analyse av kriminalitet og økonomiske forhold, kan man undersøke hvordan fattigdom og lav inntekt kan påvirke kriminalitetsnivået. Sekundær Hypotese: Økt andel fattige øker vinningskriminaliteten. I grafen under ser vi utvikling i variabelen fra 2003 til 2021. Vi ser at det er en svak positiv trend som startet rundt 2012. Andel med lavinntekt har et gjennomsnitt på 8.7%, standardavvik på 0.86 og en topp i 2019-20 hvor andelen er på ca. 10%. I datasett og regresjonsanalyse gjort i STATA er variabelen behandlet med variabelnavnet *lavinntekt*.

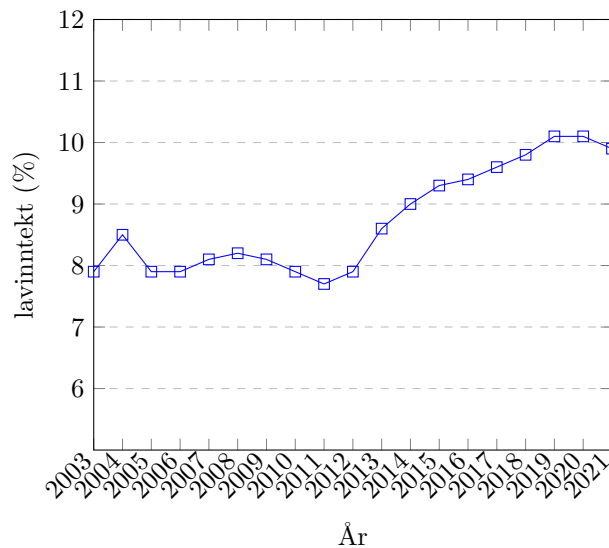


Figure 5: Prosent av befolkning med lavinntekt (2003-2021)

3.2.6 Prosentandelen av innvandrere i befolkningen

Vi har hentet kvantitative data fra Statistisk sentralbyrå som viser innvandring hvert kvartal i hvert år fra 2003 til 2021. Disse dataene gir et bilde av innvandringsmønstre i Norge over en periode på nesten to tiår. Merkelappen “innvandrere” er ikke kun innbyggere som er født utenfor Norge. Statistikken omfatter personer som har to utenlandsfødte foreldre, eller mer presist: Personer som verken har foreldre eller besteforeldre som er født i Norge. SSB har også behandlet dataen slik at personvern opprettholdes [6]. Siden dataene er kvartalsvise, må det interpoleres for å oppnå et årlig mål på innvandring som kan sammenlignes med andre variabler i analysen, slik som kriminalitetsnivåer. For å normalisere data for innvandring og gjøre dem lettere å tolke i forbindelse med vår analyse, har vi delt det totale antallet innvandrere på Norges befolkning hvert år i vår tidsperiode. Dette gir oss en relativ måling av innvandring som uttrykker innvandrere som en andel av den totale befolkningen. Denne justeringen gjør det mulig å sammenligne innvandringens påvirkning på kriminalitet uavhengig av endringer i befolkningsstørrelse over tid. Vår sekundære hypotese er at økt innvandring fører til økt kriminalitet. For å undersøke denne hypotesen, vil vi inkludere innvandrere som en andel av befolkningen som en uavhengig variabel i vår regresjonsanalyse. Dette vil tillate oss å evaluere om det er en statistisk signifikant sammenheng mellom innvandring og vinningskriminalitet i Norge. Under ser vi at andel innvandrere i Norges befolkning. Det er her en klar økning fra 2003 til 2021 med et gjennomsnitt på ca. 10.5%, standardavvik på 3.16. Topppunktet er i 2021 og er på 14.8%.

I datasett og regresjonsanalyse gjort i STATA er variabelen behandlet med variabelnavnet *Prosent_innvandrere*.

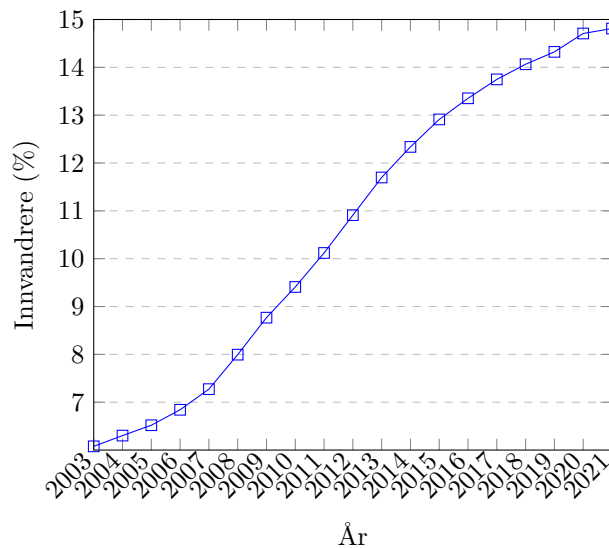


Figure 6: Prosent innvandrere av befolkningen (2003-2021)

3.2.7 Andel oppklarte eiendomstyveri

Oppklaringsprosent er prosentandel av anmeldte lovbrudd som blir klassifisert som oppklart. Data er hentet fra SSB og viser anmeldte eiendomstyveri med status oppklart og uoppklart fra 2003 til 2021 [7]. Ved å dele oppklarte lovbrudd på antall lovbrudd får vi prosentandel av eiendomstyveri som får status oppklart. merk at oppklart inneholder tilfeller der anmeldelsen ble henlagt fordi siktede ikke er strafferettslig ansvarlig, anmeldelsen er overført til konfliktråd, anmeldelse har ført til forelegg, der anmeldelse har ført til tiltale i tillegg til, påtaleunntatelse og “andre oppklarte anmeldelser” (SSB). Under ser vi grafisk utviklingen i oppklarte eiendomstyverier. Vi ser at det svinger en god del med flere lokale topp og bunnpunkter. Gjennomsnittlig utvikling er ca.18.8%, standardavvik på ca. 1.3. Toppunktet finner vi i 2016 på 21.21%.

I datasett og regresjonsanalyse gjort i STATA er variabelen behandlet med variabelnavnet *oppklaring*.

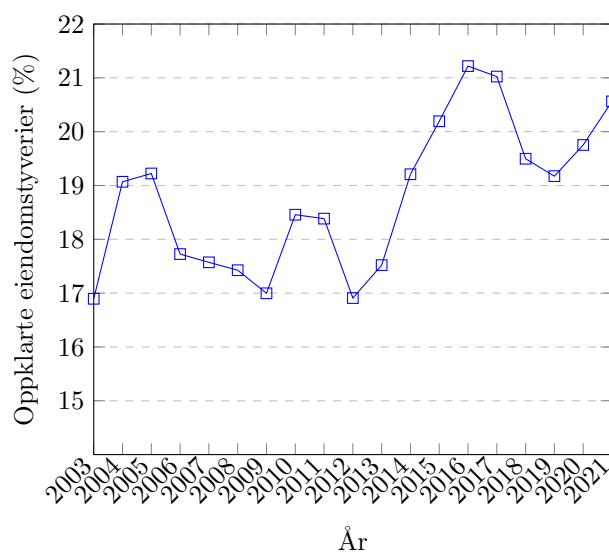


Figure 7: Prosent oppklarte eiendomstyverier (2003-2021)

3.3 Korrelasjon og korrelasjonsmatrise

I økonometri refererer begrepet korrelasjon til et statistisk mål som uttrykker styrken til sammenhengen mellom to eller flere variabler. Korrelasjonskoeffisienten, ofte betegnet som "r", har en verdi mellom 0-1, hvor 0 uttrykker ingen sammenheng og 1 antyder perfekt sammenheng. Når man ser på korrelasjonskoeffisienten mellom flere variabler, bruker man gjerne korrelasjonsmatrise. Dette er en tabell hvor hver celle fylles av korrelasjonskoeffisienten mellom to variabler. Korrelasjonsmatriser anvendes til å finne ut om hvorvidt sammenhengen mellom de ulike variablene er sterke eller svake. På denne måten kan man luke ut de variablene som ikke egner seg i regresjonsmodellen, og bygge modellen basert på denne informasjonen. Korrelasjonsmatriser kan også forutsi multikollinearitet i MLR. Ved sterk korrelasjonskoeffisient blant uavhengige variabler kan man oppdage problemer knyttet til multikollinearitet. Dog er det viktig å nevne at høy korrelasjon ikke nødvendigvis innebærer kausalitet. Dette er fordi korrelasjonskoeffisienten ikke forteller oss om x fører til y, y fører til x eller både x og y påvirker hverandre. I tillegg kan variablene være påvirket av andre faktorer som ikke blir tatt opp i modellen, og noe kan være ren tilfeldighet (Woolridge, 2012, s. 739)

Matematisk uttrykkes dette som: $x_i, y_i = \text{datamaterialet, to variabler}$

$\bar{x} = \text{gjennomsnittet av } x_i$

$\bar{y} = \text{gjennomsnittet av } y_i$

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Table 1: Korrelasjonsmatrise

	Vinningskrim.	År	Arbeidsled.	Langtidsled.	Lav inntekt	BNP	Oppklaringsrate
Vinningskrim.	1.0000						
År	-0.9653	1.0000					
Arbeidsled.	-0.4071	0.3997	1.0000				
Langtidsled.	-0.7552	0.7661	0.4826	1.0000			
Lav inntekt	-0.9045	0.8725	0.5398	0.6456	1.0000		
BNP	0.0735	-0.1448	0.0543	0.1045	-0.0567	1.0000	
Oppklaringsrate	-0.7493	0.6498	0.7334	0.7350	0.7491	0.2371	1.0000

Tabellen over viser hvor mye hver variabel korrelerer med de andre variablene. Her ser vi at mange variabler er sterkt korrelert med år-variabelen, som viser hvor langt ut i den bestemte perioden 2003-2021 vi befinner oss. Det er stort sett endring i BNP og naturligvis vinningskriminalitet som korrelerer positivt med de andre variablene. Også prosent innvandrere i befolkningen og andel med lavinntekt er sterkt korrelert. Andel med lavinntekt er også sterkt korrelert med prosent av innvandrere i befolkningen ettersom innvandrere generelt har lavere inntekt. Grunnene til dette kan være mange. Noen årsaker kan være mangelfull utdanning, språkproblemer, tradisjonelle kjønnsroller, arbeidslivsdiskriminering og kulturforskjeller. Det faktum at en økning i innvanderandelen og andelen med lavinntekt korrelerer positivt med nesten 90% er derfor ikke veldig overraskende.

4 Metode

For å svare på vår problemstilling tar vi i bruk regresjonsanalyse som verktøy. I følgende kapittel vil vi først forklare viktige begreper og metoder innenfor statistikk og økonometri. Disse metodene og begrepene vil brukes for å både analysere dataen og diskutere funn senere i oppgaven. Videre vil vi skrive litt om hva hypotesetesting innebærer og hvorfor vi bruker det. I slutten av kapitlet vil vi presentere de ulike modellene våre, og forklare fremgangsmåten vi har brukt for å justere for den tidligere nevnte negative trenden i antall årlige anmeldte vinningslovbrudd.

4.1 Regresjon

”Regresjonsanalyse er statistiske analysemetoder for å beskrive sammenhengen mellom én eller flere uavhengige variabler og en avhengig variabel. Det gjøres ved å finne frem til en funksjon, det vil si et tilnærmet uttrykk eller en forenklet matematisk beskrivelse av den virkelige sammenhengen.” (Braut & Dahlum, 2021).

Når man skal utføre en statistisk analyse er det flere mulige metoder som kan benyttes avhengig av datamaterialets egenskaper. Eksempelvis lineær regresjon, multippel regresjon, logistisk regresjon og flere andre. Typen regresjonsmetode som brukes vil være avhengig av fordelingen av dataene, forutsetningene som ligger til grunn for hver metode, og arten av den avhengige variabelen (kontinuerlig, binær eller ordinær). Formålet med en regresjonsanalyse er å forstå og måle hvordan variablene påvirker hverandre, samt å kunne forutsi hvordan endringer i de uavhengige variablene vil påvirke sammenhengen mellom den avhengige variabelen (også kjent som respons- eller resultatvariabel) og en eller flere uavhengige variabler (også kjent som forklarende variabler eller inngangsvariabler). Denne informasjonen kan man bruke til å bedre forstå sammenhengen mellom makroøkonomiske størrelser og andre statistiske målepunkter.

Regresjonsanalyse brukes ofte til å avgjøre betydningen og retningen av sammenhengen mellom avhengige og uavhengige variabler. Koeffisientene i en regresjonsanalyse forteller oss noe om størrelsen på variablene og hvorvidt de har en positiv eller negativ innvirkning på utfallsvariabelen. Forskere som bruker estimerte koeffisienter som referanse, kan forutsi utfall ved å tilpasse en regresjonsmodell til dataen som er tilgjengelig. Ved å bruke regresjonsanalyse kan man evaluere virkningen av ulike faktorer på en avhengig variabel, noe som kan hjelpe forskere med å konstruere modeller for å forutsi eller forstå visse fenomener basert på data.

Ved bruk av multippel regresjonsanalyse, kan forskere inkludere en rekke uavhengige variabler for å vurdere hvor mye hver enkelt inngangsvariabel påvirker den avhengige variabelen. Ved denne type regresjonsanalyse kan man også kontrollere for variabler som motsier resultatene i analysen. Det er også vanlig å teste hypoteser ved bruk av regresjonsanalyse. Mer spesifikt ønsker man å se om en sammenheng mellom to uavhengige variabler er statistisk signifikant, eller om en bestemt inngangsvariabel har en betydelig innvirkning på resultatvariabelen. Ofte analyseres flere statistiske indikatorer, inkludert R^2 , p-verdi, og konfidensintervaller for koeffisientene, for å vurdere om resultatene er statistisk signifikante og generaliserbare til en større befolkning.

4.2 Statistikkmetode: Minste kvadraters metode

OLS, ”Ordinary Least Squares” eller minste kvadraters metode, er en statistisk tilnærming som simulerer forholdet mellom en avhengig variabel og en eller flere uavhengige variabler i en lineær regresjon. Den anvendes ved å minimere summen av de kvadrerte residualene mellom de observerte verdiene og de predikerte verdiene. På denne måten kan man estimere de ukjente parameterne til en lineær regresjonsmodell. I en bestemt populasjon kan forholdet mellom de forklarte og forklarende variablene uttrykkes

som følger: $y = \beta_0 + \beta_1 x_1 + \varepsilon$

Ved hjelp av et utvalg fra den aktuelle populasjonen forsøker man å estimere størrelsen på populasjonsparemetrene ved hjelp av denne metoden. Dette gjøres under forutsetning av at det er en lineær sammenheng mellom de aktuelle variablene. En avhengig variabel eller forklart variabel (y) og en eller flere uavhengige variabler eller forklaringsvariabler (x_i) utgjør en regresjon. I vår analyse er de avhengige og uavhengige variablene henholdsvis vinningskriminalitet og arbeidsledighet, langtidsledighet, BNP, prosent av innvandrere, lavinntekt og oppklaringsprosent.

Den enkle lineære regresjonslinjen (SLR, simple linear regression), som antas å holde populasjonen som vurderes, er definert av ligningen over [1]) (Wooldridge, 2020). Helningsparameteren, eller 1, representerer mengden av endringen i y når x varierer med én enhet, mens avskjæringsparameteren, eller 0, representerer konstanten. De uobserverte variablene, eller de andre elementene enn x som har innvirkning på y , er representert av variabelen ε (Wooldridge, 2019a, s. 21).

Vi kan estimere disse variablene ved bruk av OLS. Dette oppnås ved å redusere variansen i utvalget, eller summen av kvadrerte avvik (SSR, sum of squared residuals). Avstanden mellom regresjonslinjen og datapunktene utgjør residualene (i), som er definert som forskjellen mellom hver observasjon i prøven og verdien regresjonslinjen predikerer. For å si det på en annen måte, kan vi bruke OLS til å lage en lineær linje som estimerer den forventede effekten av en endring i en bestemt uavhengig variabel, x_i , på den avhengige variabelen, y . Dette gjøres ved å undersøke verdien av hver observasjon i et populasjonsutvalg. Med denne kunnskapen kan vi introdusere den estimerte modellen (SRF), også kjent som OLS-regresjonslinjen (Wooldridge, 2012, s. 30).

Dette er en standardisert måte å presentere resultater fra en regresjonsanalyse på, som inkluderer en del av informasjonen om modellspesifikasjoner, dataene som er brukt, variable inkludert, samt estimerer for koeffisientene, deres standardfeil og eventuelle signifikanstester. SRF brukes for å gjøre sammenligninger mellom forskning enklere og for å gi en tydelig oppsummering av variablene som ble brukt og hvordan modellen ble satt sammen.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

Variablene er estimerer av de reelle populasjonsparemetrene introdusert i ligningen. Noen forutsetninger må være sanne for at disse estimatorene – inkludert OLS – skal være objektive. ”Gauss Markov-teoremet” refererer til alle disse forutsetningene, som i dette tilfellet gjelder enkel regresjon (Wooldridge, 2012, s. 83-88).

For at de estimerte parameterne i en OLS-regresjon skal kvalifisere som Best Linear Unbiased Estimators (BLUEs), må de oppfylle de fem første Gauss-Markov-kriteriene. Disse estimatene regnes som ”Best” hvis de har den laveste variansen blant alle forventningsjusterte estimatorer, ”Lineær” hvis de er lineære kombinasjoner av den uavhengige variabelen, og ”Unbiased” hvis de samsvarer med de sanne parameterne i gjennomsnitt. Når disse betingelsene er oppfylt, estimerer de estimerte parameterne effektivt de faktiske parameterne i regresjonsanalysen.

MLR.1 Tilfeldig utvalg Denne antagelsen sier at dataene er et tilfeldig utvalg fra populasjonen. Det betyr at observasjonene er uavhengige av hverandre og representerer populasjonen vi ønsker å generalisere funnene våre til (Wooldridge, 2012, s. 83-88).

II. MLR.2: Linearitet i parametre Sammenhengen mellom de uavhengige variablene x_i og den avhengige variabelen y er lineær og kan skrives som en rett linje med formelen $Y = \beta_0 + \beta_1 x_1 \dots + u$ (Wooldridge, 2012, s. 83-88).

III. MLR.3 Nok variasjon i x , ingen multikollinearitet: Den uavhengige variablene x_i har nok variasjon,

som indikerer at de respektive observasjonene av de uavhengige variablene er av ulik verdi. Det er heller ingen sterke lineære sammenhenger mellom de uavhengige variablene. I vårt tilfelle vil det for eksempel bety at andelen nordmenn med lavinntekt ikke blir i høy grad forklart av arbeidsledighetsraten (Wooldridge, 2012, s. 83-88).

IV. MLR.4: Null betinget gjennomsnittsverdi $E(u | x_1, x_2, \dots, x_k) = 0$ Feilledet har en forventet gjennomsnittsverdi av null, gitt hvilken som helst annen verdi av de uavhengige variablene. Korrelasjon mellom feiltermen og de uavhengige variablene kan finne sted, men det skal ikke være en sammenheng som er konsekvent og systematisk (Wooldridge, 2012, s. 83-88).

V. MLR 5: Homoskedastisitet: Variansen til feilledet (residuale) er konstant for alle observasjoner i en regresjonsmodell. Forskjellen mellom de observerte verdiene av den avhengige variabelen (Y) og de forventede verdiene (basert på regresjonsmodellen) er ikke relatert til de uavhengige variablene. Videre betyr det at variansen til feilene er konstant, noe som betyr at spredningen av feilene er lik over hele spekteret av de uavhengige variablene (Wooldridge, 2012, s. 83-88).

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

4.3 Goodness of fit

Goodness of fit er en matematisk modell som sikter på å forklare og redegjøre for forskjellene mellom observerte data og estimerte data. I utgangspunktet er goodness of fit brukt til å analysere hvorvidt utvalget av data passer til en populasjon som har en normalfordeling. Med dette kan man se om utvalgsdataen er gyldig, og hvorvidt den gjenspeiler anomalier som vil være til stede i hele populasjonen. R^2 anslår hvor mye av variasjonen i den avhengige variabelen som kan forklares av hver av modellens uavhengige variabler. Hovedessensen er at hver uavhengig variabel i modellen bidrar til forklaringen av variasjonen i den avhengige variabelen. R-kvadrert måles fra 0% til 100%. Når R-kvadrert er 100%, betyr det at modellen forklarer all variasjonen i den avhengige variabelen. En verdi på 0% indikerer at dataen ikke er forklart av modellen.

Justert R-kvadrert angir prosentandelen av variasjon som kan forklares av de uavhengige variablene, noe som har innvirkning på forklaringen av den avhengige variabelen. Forskjellen mellom R-kvadrert og justert R-kvadrert at førstnevnte tar hensyn til alle uavhengige variabler, mens justert R-kvadrert kun tar hensyn til de uavhengige variablene som faktisk har innvirkning på modellen. Når man bygger en regresjonsanalyse vil R^2 fortsette å stige med hver uavhengig variabel lagt til, men justerte R^2 vil kun øke når den uavhengige variabelen som blir lagt til faktisk påvirker den avhengige variabelen. R^2 vil stige selv om en variabel ikke er signifikant, men justert R^2 vil reduseres når dette skjer. Vi har valgt å benytte R^2 , ettersom vi bruker modeller med varierende antall uavhengige variabler.

4.4 Variansinflasjonsfaktor (VIF)

VIF er en måleenhet på mengden av multikollinearitet i regresjonsanalysen. Multikollinearitet oppstår når det er en sammenheng mellom én eller flere av de uavhengige variablene i regresjonsanalysen. Dette kan negativt påvirke analysen. Det er derfor nyttig å bruke VIF for å estimere hvor mye av variansen til koeffisientene er oppblåst som et resultat av multikollinearitet. Formelen for VIF er:

$$\frac{1}{1-R_i^2}$$

R_i^2 symboliserer hvor stor forklaringskraft den aktuelle variabelen har på de andre uavhengige variablene. Dette kan man finne manuelt ved å kjøre en regresjon med den aktuelle variabelen som avhengig/forklart variabel. Formelen viser at VIF blir lavere hvis R_i^2 er nærmere 0.

Multikollinearitet kan danne et problem i regresjonsanalysen dersom variablene i stor grad påvirker hverandre. De uavhengige variablene vil derfor ikke lenger kunne kvalifiseres som uavhengige. Selv om multikollinearitet ikke reduserer den aggregerte forklaringskraften til modellen, så kan den produsere estimater av koeffisienter som ikke er statistisk signifikante. Det kan da føre til en form for dobbel medregning av resultatene. En høy grad av multikollinearitet vil gjøre det vanskeligere å estimere forholdet mellom hver uavhengig variabel og den avhengige variabelen. Vi kan derfor si at når to eller flere uavhengige variabler er nært korrelert eller måler omtrent det samme, så kan den underliggende effekten de måler bli regnet med flere ganger. Det er dette tester av variansinflasjonsfaktor kan hjelpe oss med å avdekke. I testene så har vi generelle tommelfingerregler når det gjelder verdiene vi får. En VIF på 1 forteller oss at variablene ikke er korrelert, En VIF mellom 1 og 5 indikerer at variablene er moderat korrelert og en VIF på mer enn 5 betegnes som problematisk for å estimere koeffisienten (Wooldridge, 2012, s. 98)

4.5 Hypotesetesting og signifikans

For å finne ut om funn i statistiske analyser er signifikante, kan man bruke hypotesetesting. Når vi kjører ulike regresjoner på den avhengige variabelen vinningskriminalitet får vi oppgitt verdiene på koeffisientene til de forskjellige uavhengige variablene. Disse verdiene er imidlertid bare estimater basert på datagrunnlag og modellestimeringer som alltid har visse begrensninger. Med andre ord kan vi aldri vite de faktiske forholdene mellom y-leddet og x-leddene med 100% sikkerhet. Derfor formulerer vi ulike hypoteser om våre variabler, og tester disse hypotesene ut ved hjelp av statistisk inferens. (Woolridge, 2012, s. 121).

Denne oppgaven sikter på å undersøke forholdet mellom arbeidsledighet og vinningskriminalitet. Dermed er det hensiktsmessig å utføre en hypotesetest på koeffisienten til arbeidsledighet, som altså er vår interessevariabel. Vi formulerer først nullhypotesen og alternativhypotesen. Som regel er det nullhypotesen man først og fremst ønsker å utfordre/teste (Woolridge, 2012, s. 122). I vårt tilfelle vil nullhypotesen hevde at det ikke finnes en sammenheng mellom vinningskriminalitet og arbeidsledighet. Dette betyr at nullhypotesen uttrykker at koeffisienten/stigningstallet vil tilsvare 0, og at ledighet dermed ikke har noen effekt på vinningskriminalitet når man har kontrollert for påvirkningen fra de andre variablene nevnt i kapittel 3. Det er altså den partielle effekten på den avhengige variabelen man ønsker å undersøke (Woolridge, 2012, s. 122). Alternativhypotesen, ofte referert til som forskningshypotesen, vil hevde det motsatte, nemlig at det eksisterer en sammenheng. I vårt tilfelle vil denne påståtte sammenhengen være positiv. Hypotesene vil altså være betegnet på følgende måte:

$$H_0: \beta_1 = 0 \quad H_A: \beta_1 > 0$$

Vi tar utgangspunkt i nullhypotesen, og ønsker å finne ut om vi kan beholde den eller forkaste den til fordel for alternativhypotesen. Siden vi tester kun én populasjonsparameter, bruker vi den såkalte t-testen, som baserer seg på t-distribusjonen. Testobservatoren/t-statistikken TS har følgende formel:

$$TS = (\hat{\beta}_1 - \beta_1) / SE(\hat{\beta}_1)$$

$\hat{\beta}_1$ er modellens estimerte koeffisient for arbeidsledigheten β_1 vil tilsvare 0, siden testobservatoren tar utgangspunkt i nullhypotesen, hvor $\beta_1 = 0$. $SE(\hat{\beta}_1)$ henviser til standardfeilen, som forteller hvor mye vår modells estimat kan variere fra de faktiske verdiene, altså feilmarginen. En relativt høy standardfeilverdi impliserer ofte at variabelen ikke har en isolert signifikant effekt. I hypotesetester velger man et signifikansnivå. Dette betyr at man setter en grense på hvor sannsynlig det er at funnene kun er tilfeldige, altså at de ikke forteller noe om det faktiske forholdet man prøver å undersøke. t-testene vi kommer til å utføre senere i oppgaven vil undersøke om arbeidsledighet har en positiv effekt på

vinningskriminalitet. Dermed vil vi utføre høyrehale-tester, som innebærer at forkastningsområdet er ytterst til høyre i t-distribusjonen. For å bestemme om man skal forkaste sammenligner man TS, som regnes ut ved å bruke den tidligere nevnte formelen med den kritiske verdien vi får oppgitt. Sistnevnte tilsvarer grensen/maksimumsverdien TS kan ha for at man ikke skal forkaste nullhypotesen, og bestemmes av antall frihetsgrader og det valgte signifikansnivået. Frihetsgraden, som kort forklart er det konkrete antallet uavhengige verdier som kan estimeres i en statistisk modell, tilsvarer antallet observasjoner (n) minus antallet parametre (k) minus 1: $df = n - k - 1$ (Woolridge, 2012, s. 121). Hvis TS er større enn kritisk verdi er man innenfor forkastningsområdet, og man forkaster nullhypotesen. Da konkluderer man med at den aktuelle variabelen er statistisk signifikant gitt det bestemte signifikansnivået (Woolridge, 2012, s. 129). Alternativt kan man sammenligne det valgte signifikansnivået med TS sin tilhørende p-verdi. En p-verdi er et tall mellom 0 og 1, og betegner sannsynligheten for å finne en teststatistikk som er mer ekstremt enn TS eller tilsvarende TS, gitt at H_0 stemmer. Hvis p-verdien er lavere enn signifikansnivået blir altså nullhypotesen forkastet. En fordel med denne framgangsmåten er at man kan finne ut akkurat hvor lavt signifikansnivået kan være for at man fortsatt kan forkaste H_0 (Woolridge, 2012, s. 133).

Ved hjelp av en F-test kan man sammenligne to ulike modeller, og dermed undersøke hvor godt de respektivt kan forklare variasjonene i den avhengige variabelen. Siden vårt datagrunnlag, og følgelig våre modeller, inneholder en rekke kontrollvariabler utover interessevariabelen arbeidsledighet, kan det være hensiktsmessig å undersøke hvorvidt kontrollvariablene alene har en samlet effekt på vinningskriminalitet som er signifikant. Denne framgangsmåten kalles multippel hypotesetesting (Woolridge, 2012, s. 144). Nullhypotesen vil hevde at ingen av de øvrige variablene har en signifikant påvirkning på vinningskriminalitet når man har kontrollert for arbeidsledighet. Alternativhypotesen vil hevde det motsatte, nemlig at en eller flere av kontrollvariablene har en koeffisient som ikke tilsvarer 0 i virkeligheten. Hypotesene vil altså se slik ut:

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

$$H_A : \neq H_0$$

For å kunne regne ut testobservatoren/F-statistikken manuelt kjører man en regresjon med samtlige variabler og deretter en regresjon med interessevariabel som eneste uavhengige variabel. Disse kalles henholdsvis den ikke-restriktive og den restriktive modellen. Testobservatoren i dette tilfellet vil ha følgende formel:

$$F = (R^2_{unrestricted} - R^2_{restricted}) / (1 - R^2_{unrestricted}) \times (n - k - 1) / q$$

R^2 , også kalt forklaringskraft, er definert tidligere (se kapittel 3). Naturligvis vil forklaringskraften variere mellom den restriktive og den ikke-restriktive modellen. $n-k-1$ er som tidligere forklart frihetsgraden. Merk at det blir frihetsgraden til modellen som inneholder alle kontrollvariablene. q betegner antall restriksjoner, altså antallet variabler som blir utelatt i den restriktive modellen (Woolridge, 2012, s. 150). For å kunne avgjøre om vi skal forkaste eller beholde nullhypotesen, må vi regne ut den kritiske verdien for F-distribusjonen. Frihetsgrad til telleren blir antallet restriksjoner og frihetsgrad til nevneren blir $n - k - 1$. Som i en t-test vil en F-statistikk som er større enn den kritiske verdien bety at man bør forkaste nullhypotesen. Hvis det motsatte er tilfelle, kan det være aktuelt å utelate kontrollvariablene fra modellen (Woolridge, 2012, s. 148). Man kan også bruke p-verdi-tilnærmingen for å finne ut lavest mulige signifikansnivå for at H_A fortsatt stemmer.

Vi kommer tilbake til hypotesetesting i neste kapittel. Da vil vi bruke verdier vi får oppgitt i STATA for å kunne si noe om signifikansen til de uavhengige variablene vi har valgt å bruke. Dette gjør vi for å styrke diskusjonen og analysen av resultatene fra regresjonene.

4.6 Modellene

Den første regresjonen vi kommer til å kjøre er en enkel lineær regresjon hvor arbeidsledighet er den eneste uavhengige variabelen. Grunnen til at vi begynner med en enkel modell er for å understreke nødvendigheten av å trendjustere modellen og/eller inkludere supplerende kontrollvariabler. Dette vil bli tydeligere i neste kapittel når vi diskuterer resultater. Vår SLR-modell vil ha følgende matematiske form:

$$Vinningskriminalitet = \beta_0 + \beta_1 \text{arbeidsledighet} + u$$

Neste modell er hovedmodellen i denne oppgaven. Dette er en multippel lineær regresjonsmodell hvor alle tidligere nevnte kontrollvariabler er inkludert. I tillegg har vi valgt å legge til en ekstra variabel, som er antall år, hvor år 2003 = 1, år 2004 = 2, osv. Siden antall anmeldte lovbrudd, som nevnt, har gått ned hvert år, vil inkludering av sistnevnte variabel til dels kontrollere for den negative årlige utviklingen. Helningskoeffisienten til År vil altså (delvis) vise hvilken sammenheng det er mellom omfanget av vinningskriminalitet og hvor langt unna vi tidsmessig befinner oss fra år 2003. År-variabelen fanger til en viss grad opp det faktum at vinningslovbrudd har minket med årene, antakeligvis uavhengig av makroøkonomiske og demografiske utviklinger. Da vil forholdet mellom vinningslovbrudd og variablene vi har valgt ut trolig bli bedre uttrykt. Hovedmodellen tar følgende form:

$$Vinningskriminalitet = \beta_0 + \beta_1 \text{arbeidsledighet} + \beta_2 \text{langtidsledighet} + \beta_3 \text{BNP} + \beta_4 \text{Innvandrerandel} + \beta_5 \text{lavinntekt} + \beta_6 \text{oppklaringsrate} + \beta_7 \text{år} + u$$

I tillegg til hovedmodellen, har vi valgt å også ta i bruk en alternativ modell. Modellen har en log-level form. Her bruker vi attpåtil en alternativ to-steps metode for å justere for trend. Det gjøres på følgende måte: Først settes det opp en modell med vinningskriminalitet og år som variabler, dette for å fange og isolere den iboende trenden i vinningskriminalitet over tid. Ved å lagre restverdiene fra denne modellen som en ny variabel, kan man isolere den delen av variasjonen i vinningskriminalitet som ikke kan tilskrives den generelle tidsutviklingen alene. Dette gjør det enklere å identifisere og analysere andre faktorer som kan påvirke vinningskriminalitet, som for eksempel arbeidsledighet. Deretter benyttes den naturlige logaritmen av residualene som en variabel i den økonomiske modellen. Dette gjøres av flere årsaker: For det første kan det bidra til å normalisere fordelingen av residualene, noe som gjør at modellen er mindre følsom for ekstreme verdier og mer solid. For det andre kan det hjelpe til med å tolke effektene av de uavhengige variablene mer nøyaktig, ettersom logaritmiske transformasjoner ofte kan gjøre at resultater blir mer intuitive og lettere å tolke i form av relative endringer. For å takle negative observasjoner i residualene, kan man benytte en konstant som legges til hver observasjon før logaritme-transformasjonen. Dette gjøres for å sikre at alle residualverdiene er positive før man tar logaritmen. Konstanten som legges til, bør være litt større enn den absolutte verdien av den minste negative residualen, slik at alle observasjoner blir positive. Målet med denne er å bevare de relative forholdene mellom observasjonene mens man håndterer de negative verdiene. Den logaritmiske, trendjusterte regresjonsmodellen vil ha følgende form:

$$\log(\text{residual}) = \beta_0 + \beta_1 \text{arbld} + \beta_2 \text{langled} + \beta_3 \text{BNP} + \beta_4 \text{Innvandrerandel} + \beta_5 \text{lavinntekt} + \beta_6 \text{oppklaringsrate} + u$$

Her har vi også inkludert alle kontrollvariabler, blant annet fordi det muliggjør en sammenligning av resultatene i denne supplerende modellen med resultatene i vår hovedmodell.

5 Resultater

Vi skal nå presentere resultater fra regresjonsanalyser gjort i STATA. STATA er en programvare som automatisk utfører regresjoner av våre modeller. Det kan også brukes til å gjøre andre statistiske tester. Datamaterialet som er brukt er lagt til i appendix i tillegg til “do-filen”. I tillegg vil vi tolke og diskutere hva de forskjellige koeffisientene vi får oppgitt kan si om den faktiske sammenhengen mellom vinningskriminalitet og de andre størrelsene, først og fremst arbeidsledighet.

5.1 Enkel lineær regresjon av vinningskriminalitet og arbeidsledighet

Nå skal vi presentere resultatet av vår enkle lineære modell med vinningskriminalitet og arbeidsledighet. Formålet med denne regresjonen er å se på arbeidsledighetens direkte isolerte effekt på vinningskriminalitet.

$$\text{Vinningskriminalitet} = \beta_0 + \beta_1 \text{arbl}d + u$$

Table 2: Enkel regresjon med vinningskriminalitet og år

Vinningskriminalitet	Koeffisient	Std. feil.	t	P> t	[95% konf. intervall]
Arbl	-23555.25	12816.69	-1.84	0.084	-50596.11, 3485.612
Konstant	256944.2	51397.7	5.00	0.000	148504.5, 365383.8
<i>Modellstatistikk</i>					
R-kvadrert					0.1658
Justert R-kvadrert					0.1167
F(1, 17)					3.38
Sannsynlighet > F					0.0836
Antall observasjoner					19

I denne enkle modellen får vi betakoeffisient på -23555.5, som vil innebære at en prosent økning i arbeidsledigheten vil redusere antall anmeldte vinningskriminelle lovbrudd med over 23000 enheter. Vi har imidlertid en lav t-verdi på -1.84 og en noe høy p-verdi på 0.084. Dette innebærer at denne modellen er så vidt signifikant på et 10% signifikansnivå. Vi opererer som oftest med 5% signifikansnivå før vi forkaster nullhypotesen, noe vi nå ikke kan. Koeffisienten er også overraskende nok negativ, som er det motsatte av vår hypotese. Dette kan komme av at modellen ikke er justert for år og andre kontrollvariabler.

5.2 Resultater fra hovedmodellen

Nå skal presentere resultater fra hovedmodellen. Den tar for seg alle de uavhengige variablene vi ønsker å undersøke.

$$\text{Vinningskriminalitet} = \beta_0 + \beta_1 \text{År} + \beta_2 \text{arbl}d + \beta_3 \text{langled} + \beta_4 \text{BNP} + \beta_5 \text{oppklaringsrate} + \beta_6 \text{Prosent_innvandring} + \beta_7 \text{lavinntekt} + u$$

Table 3: Regresjonsresultater av multippel lineær regresjonsmodell

Vinningskriminalitet	Koeffisient	Std. feil	t	P>—t—	[95% konf. intervall]
År	-9858.89	1471.876	-6.70	0.000	-13098.47, -6619.312
Arbld	7725.33	2747.002	2.81	0.017	1679.226, 13771.45
Langled	58.58	365.7204	0.16	0.876	-746.361, 863.5293
BNP	941.25	762.9585	1.23	0.243	-738.0084, 2620.512
Oppklaring	-8258.80	169734.7	-4.87	0.000	-1199467, -452299.6
Prosent_innvandrere	10064.62	2604.939	3.86	0.003	4331.184, 15798.05
Lavinntekt	-5075.08	2936.331	-1.73	0.112	-11537.9, 1387.74
Konstant	2.01e+07	2928997	6.85	0.000	1.36e+07, 2.65e+07
<i>Modellstatistikk</i>					
R-kvadrert					0.9912
Justert R-kvadrert					0.9855
F(7, 11)					176.11
Sannsynlighet > F					0.0000
Antall observasjoner					19

5.2.1 Trend (År)

Koeffisienten for år er -9858.89 med en standardfeil på 1471.876. T-verdien er -6.70, og p-verdien er 0.000, som er mindre enn signifikansnivået på 0.05. Dette betyr at variabelen for år er statistisk signifikant og har en negativ sammenheng med vinningskriminalitet. En økning på ett år er assosiert med en estimert reduksjon i vinningskriminalitet på 9858.89 anmeldte lovbrudd, alt annet likt.

5.2.2 Arbeidsledighet

Koeffisienten for prosent arbeidsledige er 7725.336 med en standardfeil på 2747.002. T-verdien er 2.81, og p-verdien er 0.017, som er mindre enn signifikansnivået på 0.05. Dette betyr at variabelen er statistisk signifikant og har en positiv sammenheng med vinningskriminalitet. En prosentpoengs økning i arbeidsledigheten er assosiert med en økning i vinningskriminalitet på 7725.336 enheter, alt annet likt.

Vi kan også finne elastisiteten med tanke på gjennomsnittet til variabelen med formelen:

$$\bar{x}/\bar{y} * \hat{\beta}_1 = 16344779.4/3.967895 * 7725.336 = 0.1875.$$

Vi får altså en elastisitet på 0.1875. Dette betyr at en 1% (merk: prosent, ikke prosentpoeng) økning i arbeidsledigheten er assosiert med en økning på omtrent 0.1875% i vinningskriminalitet, i gjennomsnitt og alt annet holdt konstant. Dette viser at selv om det er en positiv sammenheng mellom arbeidsledighet og vinningskriminalitet, er effekten relativt liten per prosentvise endring i arbeidsledigheten. Vår avhengige variabel er med andre ord relativt uelastisk. En stor økning i vinningskriminalitet forutsetter en stor økning i arbeidsledigheten.

5.2.3 Langtidsledighet

Koeffisienten for langtidsledighet 58.58 med en standardfeil på 365.7204. T-verdien er 0.16, og p-verdien er 0.876, som er større enn signifikansnivået på 0.05. Dette betyr at langtidsledighet ikke er statistisk signifikant i denne regresjonsmodellen. Vi kan derfor ikke konkludere med en sikker sammenheng mellom langtidsledighet og vinningskriminalitet basert på disse resultatene.

5.2.4 Endring i BNP

Koeffisienten for prosentvis endring i brutto nasjonalprodukt er 941.2518 med en standardfeil på 762.9585. T-verdien er 1.23, og p-verdien er 0.243, som er større enn signifikansnivået på 0.05. Dette betyr at variabelen *BNP* ikke er statistisk signifikant i denne regresjonsmodellen. Vi kan derfor ikke konkludere med en sikker sammenheng mellom BNP og vinningskriminalitet basert på disse resultatene.

5.2.5 Oppklaringsprosent av eiendomstyveri

Koeffisienten for oppklaringsprosenten er -825.8832 med en standardfeil på 169.7347. T-verdien er -4.87, og p-verdien er 0.000, som er mindre enn signifikansnivået på 0.05. Dette betyr at variabelen er statistisk signifikant og har en negativ sammenheng med vinningskriminalitet. En prosentpoengs økning i oppklaringsprosenten er assosiert med en estimert reduksjon på 8258.8 anmeldte lovbrudd, alt annet likt.

5.2.6 Prosent innvandrere i befolkningen

Koeffisienten for innvandrere er 10064.62 med en standardfeil på 2604.939. T-verdien er 3.86, og p-verdien er 0.003, som er mindre enn signifikansnivået på 0.05. Dette betyr at andelen innvandrere i befolkningen er statistisk signifikant og har en positiv sammenheng med vinningskriminalitet. En prosentpoengs økning i andelen innvandrere er assosiert med en økning i vinningskriminalitet på 10064.62 anmeldte lovbrudd, alt annet likt.

5.2.7 Lavinntekt

Koeffisienten for variabelen andel med lav inntekt er -5075.081 med en standardfeil på 2936.331. T-verdien er -1.73, og p-verdien er 0,112, som er større enn signifikansnivået på 0.05. Dette betyr at variabelen *lavinntekt*, ikke er statistisk signifikant i denne regresjonsmodellen. Vi kan derfor ikke konkludere med en sikker sammenheng mellom lav inntekt og vinningskriminalitet basert på disse resultatene.

5.3 Tolkning av resultater

Gjennom regresjonsanalysen på vinningskriminalitet og samtlige uavhengige variabler har vi fått følgende resultater. Vi kan bemerke oss at 3 av de uavhengige variablene ikke er av statistisk signifikans i regresjonen med et signifikansnivå på 5%. Disse variablene er henholdsvis endringer i BNP, langtidsledighet og lavinntekt. Prosentvise endringer i BNP ble tatt med i bakgrunn av en forventning om at økonomisk oppgang i Norge ville føre til at færre blir nødt til å ty til vinningskriminalitet for å dekke sine levkostnader. Tallene viser derimot ingen sterk og sikker relasjon. Andel av befolkningen som befinner seg under vår lavinntektsgrense på 60% av medianinntekt i Norge ser heller ikke ut til å ha signifikant betydning for vår avhengige variabel. Det er intuitivt å tenke at det er disse menneskene som er mest utsatt for å drive vinningskriminelle aktiviteter når tidene er harde. En mulig forklaring kan være Norges gode velferdsordninger som sørger for at personer i nød får dekket sine primære behov. Det gir en form for trygghet i tider med økonomisk usikkerhet og trangere kår. Vi ser heller ingen signifikant sammenheng mellom langtidsledighet og vinningskriminalitet. Vi forventet at en økning i andel langtidsledige ville føre til økning i vinningskriminelle lovbrudd, ettersom vi i utgangspunktet så for oss at personer i kontinuerlig dårlig økonomisk situasjon ville kunne bli mer desperate og dermed ty til kriminalitet. Resultater fra denne regresjonen indikerer imidlertid ikke dette. Vi kan se at prosentandel innvandrere og arbeidsledighet begge har en positiv signifikant effekt på antall anmeldte vinningslovbrudd.

Det som kanskje er litt overraskende er at en økning i andel innvandrere med ett prosentpoeng er assosiert

med en økning i anmeldte vinningslovbrudd på over 10000. Denne koeffisienten er altså markant høyere enn koeffisienten til arbeidsledighet, som er på 7725. Her er det trolig relevant å trekke inn den teoretiske modellen Winter-Ebmer og Raphael bruker i sin analyse. Modellen påpeker, som nevnt, at yrkeskriminelle overhodet ikke blir rammet av arbeidsledighet, siden de ikke er sysselsatt i utgangspunktet. SSB sin statistikk på arbeidsledighet vil heller ikke fange opp disse individene i utgangspunktet. Winter-Ebmer og Raphael understreker at kriminelle som supplerer inntekter fra forbrytelser med vanlige lønnsinntekter, trolig vil begå mer kriminalitet hvis de blir arbeidsledige. Det er derfor godt mulig at mesteparten av økningen i vinningskriminalitet som følge av høyere arbeidsledighet blir begått av mennesker i denne gruppen. Til sammenligning vil yrkeskriminelle nødvendigvis ikke begå mer vinningskriminalitet enn før. Hvis vi antar at yrkeskriminelle generelt begår store deler av vinningskriminaliteten, er det følgelig logisk at den estimerte koeffisienten til arbeidsledighet-variabelen ikke er høyere.

Vi kan også bemerke at andel med lavinntekt og andel innvandrere er sterkt korrelert. Det kan forklare noe av økningen i kriminaliteten. Imidlertid er ikke lavinntekt signifikant. Kanskje grad av økonomisk velstand har forskjellig påvirkning på vinningskriminalitet avhengig av hvilke subgrupper i samfunnet man ser på. Det burde derfor forskes mer på. Innvandring kan være et kontroversielt tema. Dermed er det viktig med en viss grad av forsiktighet i forskningen fordi ufullstendig forskning lett kan misbrukes til å for eksempel fremme visse politiske synspunkt.

5.4 Enkel lineær regresjon av vinningskriminalitet og år

I denne regresjonsanalysen utfører vi en regresjon mellom vinningskriminalitet og år. Formålet med dette er å danne grunnlaget for vår neste regresjon på residualene.

Table 4: Regresjonsresultater av vinningskriminalitet og arbeidsledighet

	Koeffisient	Std. feil.	t	P>—t—	[95% konf. intervall]
Arbld	-1466.243	4103.791	-0.36	0.726	-10165.89, 7233.404
År	-5862.786	435.3615	-13.47	0.000	-6785.711, -4939.861
Konstant	1.20e+07	869569.7	13.76	0.000	1.01e+07, 1.38e+07
<i>Model statistics</i>					
R-kvadrert	0.9324				
Justert R-kvadrert	0.9239				
F(2, 16)	110.28				
Sannsynlighet > F	0.0000				
Antall observasjoner	19				

5.5 Enkel lineær regresjon med trendjustert vinningskriminalitet

Etter at vi nå har gjennomført en regresjon med vinningskriminalitet og år, har vi nå muligheten til å lage en ny variabel kalt $\log(\text{residual})$. $\log(\text{residual})$ er den forventede logaritmiske verdien av residualene til regresjonen mellom år og vinningskriminalitet. Variabelen brukes for å justere for den klare negative årlige trenden i vinningskriminalitet. $\log(\text{residual}) = \beta_0 + \beta_1 \text{arbld} + u$

I regresjonsanalysen av arbeidsledighets effekt på vinningskriminalitet ser vi at vi får koeffisienten $\beta_1=0.1755$. Koeffisienten for arbeidsledighet 1 er den forventede endringen i den logaritmisk transformerte residualen av vinningskriminalitet og år for ett prosent økning i arbeidsledighetsraten. Når arbeidsledighetsraten øker med 1 prosentpoeng, forventes den logaritmisk transformerte residualen av vinningskriminalitet og år å øke med 0.1755enheter, alt annet likt. Det er viktig å merke seg at dette ikke direkte tilsvarer en 0.1755 prosent økning i lovbrudd, siden vi jobber med log-transformerte residualer og

Table 5: Regresjonsresultater av $\log(\text{residual})$, arbeidsledighet

$\log(\text{residual})$	Koeffisient	Std. feil	t	P>—t—	[95% konf. intervall]
Arbld	0.1755635	0.366559	0.48	0.638	-0.5978084, 0.9489355
Konstant	8.402511	1.469981	5.72	0.000	5.301123, 11.5039
<i>Modellstatistikk</i>					
R-kvadrert					0.0133
Justert R-kvadrert					-0.0447
F(1, 17)					0.23
Sannsynlighet > F					0.6381
Antall observasjoner					19

ikke det faktiske antallet lovbrudd. For å få et bedre bilde av effekten på det faktiske antallet lovbrudd, kan vi utføre en invers logaritmisk transformasjon. Formelen for dette er: Endring i vinningskriminalitet = $(e(0.1755) - 1) * 100$. Når vi beregner dette, får viomtrent en 19.18% endring. Derfor, når arbeidsledighetsraten øker med 1 prosentpoeng, forventes residualen av vinningskriminalitet å øke med omtrent 19.18%, alt annet likt.

Vi ser at vi får en høy p-verdi på 0.638. Et vanlig signifikansnivå ligger som regel på 5% eller 10%. I dette tilfellet så vil signifikansnivået måtte være på 63.8% for at effekten av arbeidsledighet på vinningskriminalitet skal være av statistisk signifikans. Vi kan derfor konkludere med at arbeidsledighet for seg selv ikke er av statistisk signifikans. Vi kan også merke oss 'R-kvadrert'-verdien som ligger på 0.0133. R-kvadrert er kvadratet av residualene og forteller oss hvor mye av residualene som er forklart av modellen og hvor nære dataene ligger i forhold til regresjonslinjen. Det forteller altså hvor godt dataene passer til modellen. De potensielle verdiene R-kvadrert befinner seg et sted mellom 0 og 1, hvor en høy verdi nær 1 er sett på som en god verdi for modellen og hvor nær 100% av variasjonen i våre data er forklart av modellen. Verdien vi fikk i regresjonen forteller oss at bare 1.33% av variasjonen i dataene er forklart av modellen. Det er anbefalt på en generell basis at verdien til R-kvadrert burde ligge på minst 0.1 /10% for at modellens forklaringskraft skal kunne bli sett på som betydelig.

5.6 Resultater fra alternativ modell

Vi skal nå ta for oss resultatene for den trendjusterte modellen, i bakgrunn av vårt ønske om å rense for den klart negative trenden i årlig vinningskriminalitet som vist i figur 1. Dette skal gjøres sammen med alle våre valgte variabler. $\log(\text{residual}) = \beta_0 + \beta_1 \text{arbld} + \beta_2 \text{langed} + \beta_3 \text{BNP} + \beta_4 \text{Prosent.innvandrere} + \beta_5 \text{lavinntekt} + \beta_6 \text{oppklaring} + u$

Table 6: Regresjonsresultater av trendjustert logaritme-modell

log(residual)	Koeffisient	Std. feil	t	P>—t—	[95% konf. intervall]
Arblid	1.184828	0.3950507	3.00	0.011	0.3240859, 2.045569
Langld	0.0295248	0.0577751	0.51	0.619	-0.0963564, 0.155406
BNP	0.0187371	0.1206906	0.16	0.879	-0.2442251, 0.2816992
Lavinntekt	-0.4809826	0.4472459	-1.08	0.303	-1.455448, 0.4934825
Prosent_innvandrere	0.2754092	0.1367135	2.01	0.067	-0.0224638, 0.5732823
Oppklaringsrate	-0.8265113	0.2657928	-3.11	0.009	-1.405624, -0.2473986
Konstant	20.27465	3.046599	6.65	0.000	13.63668, 26.91262
<i>Modellstatistikk</i>					
R-kvadrert					0.6506
Justert R-kvadrert					0.4759
F(6, 12)					3.72
Sannsynlighet > F					0.0251
Antall observasjoner					19

5.6.1 Arbeidsledighet

Arbeidsledighet er en signifikant estimator for vinningskriminalitet med en koeffisient på 1.1848 og en p-verdi på 0.011. Det betyr at det er en sannsynlighet på 1.1% for at resultatene er tilfeldig. En økning på ett prosentpoeng i arbeidsledighet vil føre til en forventet økning i residualen til vinningskriminalitet på omtrent 227%, ifølge en invers logaritmisk transformasjon.

5.6.2 Langtidsledighet

Langtidsledighet har en koeffisient på 0.0295, men har en høy p-verdi på 0.619, som betyr at det er en sannsynlighet på 38.1% for at resultatene er tilfeldige. En økning på ett prosentpoeng i andelen langtidsledige er forventet å øke residualen til vinningskriminalitet med omtrent 3%, med en invers logaritmisk transformasjon.

5.6.3 Endringen i bruttonasjonalprodukt

Koeffisienten for endring i bruttonasjonalprodukt (BNP) er 0.0187. vi har imidlertid en høy p-verdi på 0.879, noe som indikerer at BNP ikke er en signifikant prediktorvariabel for vinningskriminalitet i vår modell. For å tolke effekten av BNP på vinningskriminalitet, kan vi se på endringen i BNP i prosent. En økning på ett prosentpoeng i BNP er forventet å øke residualen til vinningskriminalitet med omtrent 1.8876%, ifølge en logaritmisk transformasjon.

5.6.4 Prosentandelen av innvandrere i befolkningen

Prosentandelen av innvandrere i befolkningen har en koeffisient på 0.2754, som er statistisk signifikant med en p-verdi på 0.067. En økning på ett prosentpoeng i andelen innvandrere i befolkningen er forventet å øke residualen til vinningskriminalitet med omtrent 31%, ifølge en logaritmisk transformasjon.

5.6.5 Lavinntekt

Overordnet kan man se at andel individer med lav inntekt er negativt tilknyttet vinningskriminalitet. For prosentandelen av personer med lav inntekt er koeffisienten -0.4810. Dette indikerer at med en økning på ett prosentpoeng i andelen personer med lav inntekt i befolkningen, og med alt annet holdt konstant, vil det være en assosiert reduksjon i logaritmen til residualen av vinningskriminalitet med en koeffisient

på -0.4810. Når vi konverterer dette fra logaritmisk form, finner vi at dette vil føre til en reduksjon på omtrent $(e^{-0,4810} - 1) * 100 = -38.11\%$ i residualen av vinningskriminalitet, med alle andre variabler holdt konstant. Dette er kontrært til den tilknyttede hypotesen som antok at andelen med inntekt under 60% av medianinntekten påvirket vinningskriminalitet positivt. Vi må også bemerke at p-verdien er på 0.303. Det er altså en sannsynlighet på 30.3% for at 4 er null. Dette er langt over signifikansnivåene vi opererer med.

Grunner til at vi ender opp med et slikt resultat kan være flere faktorer: En faktor kan være at viktige variabler som har innvirkning på forholdet mellom lavinntekt og vinningskriminalitet. Dersom noen av disse utelatte variablene er korrelert med både lavinntekt og vinningskriminalitet, kan dette føre til at estimatet for lavinntekt blir skjevt. En annen faktor kan være multikollinearitet. Som forekommer når to eller flere av de uavhengige variablene i modellen er sterkt korrelert med hverandre som kan gjøre estimatene ustabile og vanskelig å tolke. Det kan også være slik at sammenhengen mellom lavinntekt og vinningskriminalitet ikke er lineær. Dette kan føre til feilaktige estimater og tolkninger av koeffisientene.

5.6.6 Oppklaringsprosent

Når det gjelder oppklaringsprosenten for anmeldte forbrytelser, er koeffisienten -0.8265. Standardfeil er 0.26579247 t-verdi på -3.11 Dette betyr at med en økning på ett prosentpoeng i oppklaringsprosenten for anmeldte forbrytelser, og med alt annet holdt konstant, vil det være en assosiert reduksjon i logaritmen til residualen av vinningskriminalitet med en koeffisient på -0.8265. Når vi konverterer dette fra logaritmisk form, finner vi at dette vil føre til en reduksjon på omtrent $(e^{-0.8265} - 1) * 100 = -55.62\%$ i residualen av vinningskriminalitet, alt annet likt.

I vår modell er koeffisientene for arbeidsledighet, oppklaringsprosent, og prosentandelen av innvandrere i befolkningen alle signifikante med en p-verdi under 0.1. Koeffisienten for arbeidsledighet har også en høy t-verdi på 3.00, mens koeffisientene for oppklaringsprosent og prosentandelen av innvandrere i befolkningen har t-verdier på henholdsvis -3.11 og 2.01. Disse t-verdiene indikerer en stor avstand mellom koeffisientene og nullhypotesen, og gir ytterligere støtte til at koeffisientene er signifikante.

Koeffisientene for langtidsledighet, BNP og lavinntekt, på den annen side, har alle høy p-verdi og lave t-verdier. Dette indikerer at disse koeffisientene ikke er signifikante i vår modell, og det er høy sannsynlighet for at koeffisientene faktisk er null.

5.7 Hypotesetesting av variabler i modellene

Ovenfor har vi oppsummert både p-verdiene og t-verdiene for de ulike stigningstallene i begge MLR-modeller. Skulle vi gjennomført hypotesetester på variablene for arbeidsledighet i MLR-modellene med et signifikansnivå på 5%, hadde vi forkastet nullhypotesen i begge tilfeller. Det som derimot ikke har blitt avdekket ennå, er om kontrollvariablene i de respektive regresjonsmodellene har en samlet signifikans. For å finne ut av dette, er vi nødt til å vite F-verdiene og/eller p-verdiene. Vi kjører derfor regresjon av hovedmodellen og skriver inn følgende kommando i STATA:

```
test langed BNP Prosent_innvandrere oppklaring lavinntekt
```

Vi får oppgitt følgende i STATA:

```
F (5, 11) = 14.53 og Prob >F = 0.0002.
```

Prob >F viser p-verdien. Vi ser altså at p-verdien er veldig lav, og vi bør forkaste en nullhypotese som hevder at kontrollvariablene ikke har en kombinert signifikant effekt på vinningskriminalitet. P-verdien uttrykker med andre ord at vi er 99.98% sikre på at det ikke er slik at samtlige kontrollvariabler egentlig

har en helningskoeffisient på 0. Det er imidlertid stor sannsynlighet for at disse verdiene ikke er helt presise. Vi har valgt å ikke inkludere år-variabelen i denne F-testen, ettersom den kun eksisterer for å forsøke å delvis fange opp negativ årlig utvikling i vinningslovbrudd. R^2 i hovedmodellen er svært høy, og dette er i stor grad takket være år-variabelen. Høy R^2 i den ikke-restriktive modellen vil gi en høy F-verdi, og følgelig lavere p-verdi. Basert på disse innsiktene, kan vi konkludere med at vi har en unøyaktig høy F-verdi i denne testen.

Når vi gjør det samme for den supplerende modellen får vi:

$$F(5, 12) = 4.38 \text{ og } \text{Prob} >F = 0.0168$$

Her har vi en høyere frihetsgrad, siden k (antall parametre) er lavere. Grunnen er at vi, som kjent, ikke har med en egen år-variabel i denne modellen. Også i denne testen får vi en relativt høy F-verdi og lav p-verdi, dog lavere enn i den første testen. P-verdien er på 0.0168. Dette betyr at vi forkaster nullhypotesen helt ned til et signifikansnivå på 1.68%. Vi kan dermed konkludere med at det er hensiktsmessig å inkludere kontrollvariablene i denne modellen.

Vi får altså resultater som tydelig indikerer forkastning av nullhypotesen i den andre F-testen. Dermed er det grunn til å tro at våre utvalgte kontrollvariabler bidrar til å generelt styrke analysen og gi et bedre bilde av sammenhengen mellom arbeidsledighet og vinningskriminalitet, til tross for det faktum at år-variabelen alene fører til en veldig høy R^2 i hovedmodellen. Dessuten har flere av kontrollvariablene i begge modeller såpass lave p-verdier at det uansett er relativt trygt å anta at de kan være med på å forklare variasjonen i vinningskriminalitet.

5.8 VIF

Det er som nevnt viktig å se på variansinflasjonsfaktoren for å avdekke eventuell multikollinearitet i modellen vår slik at effekten av de uavhengige variablene ikke blir medregnet flere ganger. Tidligere i oppgaven refererte vi til tommelfingerregler hvor en VIF mellom 1 og 5 er regnet som moderat korrelasjon og en VIF større enn 5 betegner en høy korrelasjon. I VIF-testen utført ovenfor kan vi se at langtid-sledighet, arbeidsledighet og bruttonasjonalprodukt kan kvalifiseres som moderat korrelert og derfor i liten grad lager problemer for modellen. Vi har derimot noe høyere VIF-verdier for innvandrereprosenten, lavinntekt og oppklaringsprosent på over 5. Variablenes plass i modellen kan derfor diskuteres. Ettersom verdien ikke er over 10 for noen av de uavhengige variablene velger vi derfor å beholde dem i modellen.

Table 7: Variance inflation factor (VIF)

	VIF	1/VIF
Prosent innvandrere	7.787	0.128
Lavinntekt	6.255	0.16
Oppklaring	5.672	0.176
Langled	4.188	0.239
Arbld	2.315	0.432
BNP	1.402	0.713
Gjennomsnittlig VIF	4.603	

Det viktigste å ta ut ifra VIF-testen er verdien til arbeidsledighet som ligger på 2.315 ettersom det er hovedvariabelen vi undersøker i oppgaven. 2.315 er en relativt lav verdi som betyr at effekten arbeidsledighet har på vinningskriminalitet i liten grad er forklart av de andre uavhengige variablene i modellen.

5.9 Sammenligning av resultater fra hovedmodell og trendjustert logaritmisk modell

I begge de multiple modellene vi nå har gått igjennom finner vi en betydelig positiv sammenheng mellom antall anmeldte lovbrudd av typen vinningskriminalitet og arbeidsledigheten. I begge våre enkle modeller uten kontrollvariabler får vi ingen signifikante resultater. Når vi legger til en rekke kontrollvariabler klarer vi derimot å luke ut støy i datamaterialet for å kunne se på effekten arbeidsledighet har på vinningskriminelle lovbrudd. Innvanderandel og oppklaringsprosenten er signifikante med forventet fortegn som samsvarer med sekundær-hypotesene. Imidlertid forklarer årsvariabelen det aller meste av modellen med en R-kvadrert på 0.93. Den trendjusterte modellen ser nærmere på arbeidsledighet og vinningskriminalitet i residualen uten den årlige negative trenden. Begge modeller gir statistisk signifikante, positive koeffisienter for arbeidsledighet. At begge de multiple analysene samsvarer, styrker troverdigheten til resultatet. Disse resultatene samsvarer også med tidligere resultatene av tidligere studier vi har sett på som Edmark, 2005, Altindag, 2012 og Raphael Winter-Ebmer 2001. Forskerne i disse studiene brukte datagrunnlag fra henholdsvis Sverige, store deler av Europa forøvrig, og USA.

Det faktum at vi også finner samme effekt i Norge fra 2003 til 2021 kan tyde på et generelt positivt forhold mellom arbeidsledighet og vinningskriminalitet i de fleste samfunn, til tross for relativt store variasjoner i kultur, myndighetenes rolle i økonomien, velstandsnivå og rettssystem.

5.10 Kritikk og robusthet

Det er mange mulige innfallsvinkler man kan ta når man forsker på kriminalitet. En utfordring det er nærmest uunngåelig å støte på når kriminelle forhold er forskningstema, er mørketall. Omfanget av vinningskriminalitet i Norge er nok større enn hva statistikken sier, i og med at ikke alle vinningskriminelle handlinger blir anmeldt. Dette trenger imidlertid ikke være et stort problem for vår oppgave, så lenge andelen av vinningslovbrudd som ikke blir anmeldt ikke har endret seg markant i tidsperioden vi har undersøkt. På den annen side har, som nevnt, digitale vinningslovbrudd blitt vanligere, samtidig som terskelen for å anmelde slike lovbrudd trolig er høyere enn for mer tradisjonelle vinningskriminelle handlinger. Dermed kan datagrunnlaget være upresist, og følgelig vil estimatene vi har fått bli mindre nøyaktige.

Vi har i denne analysen kun sett på Norge som helhet, som gir et ganske generelt innblikk i arbeidsledighetens påvirkning på vinningskriminalitet. Det kan derfor være gunstig i videre forskning å undersøke regionale forskjeller i kriminalitet og arbeidsledighet. Andre interessante tilpasninger kan være å sortere etter forskjellige kategorier av vinningskriminalitet eller lage en ekstra modell med voldelig kriminalitet (som mord, voldtekt etc.) som en avhengig variabel, og sammenligne modellene. Det kan også være en fordel å gjøre nye undersøkelser med en lengre tidshorisont. Samtidig er ikke dette sikkert, ettersom undersøkelser av samfunnsforhold langt tilbake i tid fort kan bli mindre relevante. Et annet forslag kan være å supplere med sosiologisk forskning hvor man sammenligner regionale eller bilaterale kultur- og holdningsforskjeller. Fordelen med en slik analyse kan være å få større innblikk i fenomener som ikke er strengt økonomiske.

Rent metodemessig kunne vi med fordel ha undersøkt potensielle ikke-lineære sammenhenger, spesielt på variabler som ikke ga signifikante resultater. Vi kunne også valgt å bruke flere variabler for å øke forklaringskraften til modellene. Siden vi bruker tidsseriedata kan en mulig feilkilde være at variabler er tilknyttet en viss latenstid før de viser effekten på vinningskriminalitet. En annen svakhet ved oppgaven er arbeidsledigheten. En bedre tilnærming kunne være å samle data på personer uten arbeid uavhengig av om de regnes som del av arbeidsstyrken for å fange opp potensielle yrkeskriminelle. Med tall på hvor

mange nordmenn som verken har jobb eller ønsker å jobbe hadde det trolig vært lettere å teste hvorvidt general strain theory også gjelder i det norske samfunnet.

I tillegg er det også usikkert i hvilken grad de øvrige MLR-forutsetningene holder i våre modeller, og følgelig hvor gyldige/presise resultatene vi har fått faktisk er. Det kan være utelatte variabler i modellene våre (for eksempel den nevnte variabelen alkoholkonsum) som påvirker både omfanget av vinningskriminalitet og en eller flere av de uavhengige variablene. Dermed er det mulig at blant annet MLR-forutsetningen om uavhengighet ikke har blitt oppfylt.

6 Konklusjon

I vår analyse har vi undersøkt sammenhengen mellom arbeidsledighet og vinningskriminalitet, både med og uten kontrollvariabler som år, langtidsledighet, endring i brutto-nasjonalprodukt, andel av befolkningen med lav inntekt, andel innvandrere i befolkningen og politiets oppklaringsrate for eiendomstyveri. Vi har funnet et statistisk signifikant positivt forhold mellom arbeidsledighet og vinningskriminalitet, både i vår hovedmodell og vår trendjusterte modell. Til tross for positiv koeffisient for arbeidsledighet er elastisiteten i forhold til gjennomsnittet uelastisk i hovedmodellen. Dette indikerer at arbeidsledighet skal øke mye før man får en relativ stor økning i vinningskriminalitet. Videre finner vi også positiv statistisk signifikant sammenheng mellom andel innvandrere, og negativ, statistisk signifikant sammenheng mellom oppklaringsraten og vinningskriminalitet. Årlige variasjoner i mørketall og forsinket effekt av arbeidsledighet på antall anmeldte vinningslovbrudd er to mulige feilkilder.

Anbefalinger til videre forskning er hovedsakelig å benytte seg av mer regionspesifikke data, dele vinningslovbrudd inn i flere underkategorier og muligens utvide modellen med flere variabler.

7 Referanseliste med kilder og statistikkbank

7.1 Kilder

- Agnew, R. & Brezina, T. (2019). General Strain Theory. In M. D. Krohn, N. Hendrix, G. Penly Hall & A.J. Lizotte (Eds.), *Handbook on Crime and Deviance* (2nd ed., pp. 145-160). Springer, Cham.
- Altindag, D.T. (2012). Crime and unemployment: Evidence from Europe. *International Review of Law and Economics*, 32(1), 145-157. <https://doi.org/10.1016/j.irle.2011.10.003>.
- Arbeids - og inkluderingsdepartementet (2023, 11. april). Arbeidsmarkedspolitikken - ansvarssområder. Regjeringen.no.
<https://www.regjeringen.no/no/tema/arbeidsliv/arbeidsmarked-og-sysselsetting/innsikt/den-norske-arbeidsmarknaden/institusjoner-og-organisering/ansvarsfordeling/id86901/>
- Braut, G.S. & Dahlum, S. (2021, 22. desember). Regresjonsanalyse. Store norske leksikon. Hentet 23. april 2023 fra <https://snl.no/regresjonsanalyse>
- Edmark, K. (2005). Unemployment and Crime: Is There a Connection? *The Scandinavian Journal of Economics*, 107(2), 353-373. <https://doi.org/10.1111/j.1467-9442.2005.00412.x>
- Lomell, Heidi Mork: vinningslovbrudd i Store norske leksikon på snl.no. Hentet 10. mai 2023 fra <https://snl.no/vinningslovbrudd>
- Meld. St. 29 (2019-2020). Politimelding: et politi for fremtiden. Justis- og beredskapsdepartementet. <https://www.regjeringen.no/no/dokumenter/meld.-st.-29-20192020/id2715224/?ch=2>
- Raphael, S & Winter-Ebmer (2001). Identifying the Effect of Unemployment on Crime. *The Journal of Law and Economics*, 44(1), 259-283. <https://doi.org/10.1086/320275>
- Statistisk sentralbyrå. (2017). *Standard for lovbruddstyper*. <https://www.ssb.no/klass/klassifikasjoner/146>
- Statistisk sentralbyrå. (2023, 26. april). *Prosessmodell for statistikkproduksjon: Kvalitet i offisiell statistikk*. <https://www.ssb.no/omssb/kvalitet-i-offisiell-statistikk/prosessmodell-for-statistikkproduksjon>
- Wooldridge, J.M. (2012). *Introductory Econometrics: A Modern Approach* (5th ed.). Cengage Learning.

7.2 Statistikkbank

- 1 Statistisk sentralbyrå, 09405: Lovbrudd etterforsket, etter statistikkvariabel, lovbruddstype, politiets avgjørelse og år, 2023. Hentet fra: <https://www.ssb.no/statbank/table/09405/>
- 2 Statistisk sentralbyrå, 07458: Sysselsetting, arbeidsledighet og utførte ukeverk, etter alder. Brudd- og sesongjustert, kvartalstall 1997K1 - 2023K1, 2023. Hentet fra: <https://www.ssb.no/statbank>
- 3 Statistisk sentralbyrå, Arbeidssøkingens lengde. Prosent av arbeidsledige. 1996-2021, 2023. Hentet fra: https://www.ssb.no/arbeid-og-lonn/sysselsetting/statistikk/arbeidskraftundersokelsen/artikler/antall-langtidsledige-pa-samme-niva-som-for-pandemien?fbclid=IwAR3Wo-6a-PQM1nJ7-M_FVWQCJ0mCIv2MJ04OQb7gILnWSfVyJSK2vUuRdJc
- 4 Statistisk sentralbyrå, 09189: Makroøkonomiske hovedstørrelser, etter makrostørrelse, år og statistikkvariabel, 2023. Hentet fra: <https://www.ssb.no/statbank/table/09189/tableViewLayout1/>
- 5 Statistisk sentralbyrå, 04204: Vedvarende lavinntekt. Treårsperiode, etter intervall (år), mottakergruppe og statistikkvariabel, Hentet fra: <https://www.ssb.no/statbank/table/04204/tableViewLayout1/?loadedQueryId=10021630&timeType=from&timeValue=1996-1998>
- 6 Statistisk sentralbyrå, 05182: Personer, etter statistikkvariabel, innvandringskategori og år, 2023. Hentet fra: <https://www.ssb.no/statbank/table/05182>
- 7 Statistisk sentralbyrå, 09405: Lovbrudd etterforsket, etter statistikkvariabel, lovbruddstype, politiets avgjørelse og år, 2023. Hentet fra: <https://www.ssb.no/statbank/table/09405/>
- 8 Statistisk sentralbyrå, 01222: Befolkning og kvartalsvise endringer, etter kvartal og statistikkvariabel, 2023. Hentet fra: <https://www.ssb.no/statbank/table/01222/>

8 Appendiks

Table 8: Blokkinformasjon fra blokkregresjon

Block	F	df	Block Residual		R2	Change in R2
			df	Pr > F		
1	232.35	1	17	0.0000	0.9318	
2	0.13	1	16	0.7255	0.9324	0.0005
3	0.07	1	15	0.7886	0.9327	0.0003
4	0.85	1	14	0.3733	0.9365	0.0038
5	7.25	1	13	0.0185	0.9592	0.0227
6	31.48	1	12	0.0001	0.9888	0.0295
7	2.99	1	11	0.1118	0.9912	0.0024

Datsett

Table 9: Data brukt i STATA

År	Vinnings- kriminalitet	BNP	arblid	populasjon	opp- klaring	lav- inntekt	lang- led	Prosent_ innvandrere
2003	225354	0.9	4.2	4562070.25	0.445771634	7.9	20	0.060775478
2004	208181	4	4.3	4587207.25	0.471624323	8.5	25.5	0.063023967
2005	194864	2.7	4.5	4618029.25	0.490455404	7.9	25.75	0.065189063
2006	189465	2.5	3.58	4655444.75	0.489385346	7.9	31.25	0.068417523
2007	182846	2.9	2.7	4699708.5	0.490566868	8.1	25.5	0.072734298
2008	177251	0.5	2.9	4761659.5	0.496652737	8.2	17.75	0.079939357
2009	189234	-1.9	3.45	4819919	0.460721545	8.1	25.25	0.087676785
2010	179068	0.8	3.95	4881879	0.4819709	7.9	31.5	0.094092049
2011	174875	1.1	3.55	4945895.25	0.48521018	7.7	32.25	0.101195026
2012	183021	2.7	3.5	5011116.75	0.473636962	7.9	28.75	0.109103824
2013	173622	1	3.95	5072270.5	0.481739603	8.6	29.25	0.116973454
2014	159274	2	3.78	5131892.25	0.499202748	9	33	0.12336775
2015	144226	1.9	4.68	5184417.25	0.513689225	9.3	32.75	0.129113836
2016	136208	1.2	4.9	5231558.25	0.520267499	9.4	37.5	0.133526182
2017	123794	2.5	4.35	5273378.25	0.505870093	9.6	37.25	0.137480561
2018	123406	0.8	4	5308668.25	0.490284967	9.8	35.5	0.140649399
2019	123835	1.1	3.9	5341773.5	0.489730116	10.1	35.25	0.143231082
2020	113004	-1.3	4.75	5374829.5	0.481454714	10.1	31.75	0.147073875
2021	104581	3.9	4.45	5401877.5	0.485506979	9.9	33.75	0.148114058

do-fil til STATA

```
clear
import excel "C:av bachelor (7).xlsx", sheet("Sheet1")
cellrange(O95:Z114) firstrowe
ssc install asdoc
reg Vinningskriminalitet År
predict uhat, residuals
gen uhat1=uhat+12398.36
gen luhat=log(uhat1)
replace Prosent_innvandrere = Prosent_innvandrere*100
regress Vinningskriminalitet arblld
regress Vinningskriminalitet År arblld langled BNP Prosent_innvandrere oppklaring lavinntekt
regress luhat arblld langled BNP Prosent_innvandrere oppklaring lavinntekt
korr Vinningskriminalitet År arblld langled BNP Prosent_innvandrere oppklaring lavinntekt
vif
test
nestreg:regress Vinningskriminalitet År arblld langled BNP Prosent_innvandrere oppklaring lavinntekt
```

