

Ola Krogh Halvorsen

## Veldedige lotterier og skattebyrde

En empirisk undersøkelse av pantelotteriets  
fordelingseffekter

Masteroppgave i Samfunnsøkonomi  
Veileder: Bjarne Strøm  
Trondheim, mars 2018

Norges teknisk-naturvitenskapelige universitet  
Institutt for samfunnsøkonomi

## Forord

Denne oppgaven markerer avslutningen på et femårig masterprogram i samfunnsøkonomi ved NTNU.

Jeg vil med dette takke for fem interessante, formende og lærerike år som samfunnsøkonomistudent ved NTNU og UCSD. En stor takk rettes også til Bjarne Strøm, for strålende veiledning, konstruktive kommentarer og inspirerende tilbakemeldinger. I en masteroppgave hvor mye blir gjort for første gang, er det godt å ha en rutinert og grundig veileder.

Denne oppgaven kunne ikke ha blitt til uten et godt datamateriale. Jeg vil derfor takke Gaute Langdal og Peter Jonas Johannesson fra Pantelotteriet og Olav Thon Gruppen for tilgang på et veldig spennende datamateriale og for god informasjon om pantelotteriet. En takk sendes også til San Diego og Marc Jacobsen ved UCSD som allerede i 2015 satte meg på tanken om å bruke data fra pantelotteriet.

Til slutt vil jeg takke alle som har gjort livet som student til en real fest. Uten førsteklasses vennskap, lærerike uker, festlige helger og opphetede samtaler over “dagens suppe”, hadde denne masteroppgaven aldri sett dagens lys.

Trondheim, mars, 2018.

Ola Krogh Halvorsen



## Sammendrag

Denne oppgavens formål er å undersøke fordelingseffektene til pantelotteriet. Pantelotteriet er et veldedig lotteri, drevet av Røde Kors, hvor man kun kan kjøpe lodd med penger man får tilbake ved retur av tomgods. Oppgaven tar utgangspunkt i et paneldatasett som følger alle pantetransaksjoner ved 2357 panteautomater fra 2012 til 2015.

Med en premieandel på 35%, innebærer pantelotteriet i likhet med andre lotterier en implisitt skatt i form av et forventet tap. Fordelingseffekten av denne implisitte skatten blir undersøkt ved kalkulering av Suits-indeks og estimering av inntektselastisiteter.

Resultatene fra begge metoder antyder at pantelotteriet er progressivt beskattende. Suits-indeksen er et deskriptivt mål på fordelingen av skattebyrde, som antar en verdi mellom 0 og 1 ved en progressiv beskatning, og en verdi mellom -1 og 0 ved en regressiv beskatning. Denne oppgaven finner en gjennomsnittlig Suits-indeks på 0.053. Estimeringer av brutto inntektselastisitet viser at konsum av pantelotteriet øker mer enn proporsjonalt med inntekten. Ved inkludering av demografiske kontrollvariabler reduseres den estimerte inntektselastisiteten betraktelig, og er ikke lenger signifikant høyere enn 1. Resultatene i denne oppgaven skiller seg fra studier av fordelingseffektene til andre lotterier, som stort sett finner at lotterier er regressivt beskattende.

Både inntektselastisiteten og Suits-indeksen viser en progressiv beskatning for samtlige utvalgsår, men beveger seg mot en mer proporsjonal beskatning gjennom utvalgsperioden. Begge målene på pantelotteriets fordelingseffekter er robuste til hvordan inntekten i befolkningen defineres.



## Abstract

This master thesis assesses the tax incidence of the Norwegian recycling-lottery. The recycling lottery is a charitable lottery, where one only can buy tickets with refundable deposits for beverage containers. Following 2357 recycling stations with the lottery function from 2012 to 2015, I construct a longitudinal dataset that is utilized in the analysis of the distributional effects of the recycling lottery.

The recycling lottery's payout percentage is 35%. The expected loss of the recycling lottery should, like other lotteries, be considered as an implicit taxation. The distributional effects of the recycling lottery is examined by the use of Suits-indexes of tax concentration and income elasticities.

The findings, by each technique, suggest a progressive taxation. Suits-indexes are a descriptive measure of tax-progressiveness, which assumes a value between 0 and 1 for a progressive taxation, and a value between -1 and 0 for a regressive taxation. Calculations of the Suits-indexes of the recycling lottery find an average Suits-index of 0.053. Estimations of the gross income elasticity finds that consumption of the recycling lottery increases more than proportionally with income. The estimated income elasticities are reduced considerably by the inclusion of demographic control variables, and yields no longer elasticities significantly higher than 1. The findings in this thesis are inconsistent with findings in prior studies of tax incidence of lotteries, which in general find regressive distributional effects.

Both measures of tax incidence find that the distributional effects are decreasingly progressive throughout the sample period. The findings are robust to different definitions of income.



# Innhold

<b>Forord</b>	<b>i</b>
<b>Sammendrag</b>	<b>iii</b>
<b>Abstract</b>	<b>v</b>
<b>1 Introduksjon og motivasjon</b>	<b>1</b>
1.1 Pantelotteriet . . . . .	2
1.2 Lotteri som beskatning . . . . .	2
1.3 Oppgavens formål . . . . .	3
1.4 Disposisjon . . . . .	3
<b>2 Teorigrunnlag og litteraturgjennomgang</b>	<b>5</b>
2.1 Hvorfor deltar vi i lotterier? . . . . .	5
2.2 Lotterier og fordelingseffekter . . . . .	14
<b>3 Metode</b>	<b>17</b>
3.1 Suits-Indekser . . . . .	17
3.2 Inntektselastisitet . . . . .	22
<b>4 Datamateriale</b>	<b>23</b>
4.1 Data fra Pantelotteriet . . . . .	23
4.2 Lotterisalg (Avhengig variabel) . . . . .	26
4.3 Inntektsvariabel . . . . .	27
4.4 Kontrollvariable i økonometrisk modell . . . . .	29
4.5 Beregnede Suits-indekser . . . . .	34
<b>5 Økonometrisk rammeverk for estimering av inntektselastisitet</b>	<b>37</b>
5.1 Økonometriske utfordringer . . . . .	39
5.2 Endogene forklaringsvariable . . . . .	39
<b>6 Modellspesifikasjoner</b>	<b>49</b>
6.1 Grunnmodellen . . . . .	49
6.2 Clustering . . . . .	51
6.3 Test av restriksjon . . . . .	51



<b>7 Resultater fra estimering av økonometrisk modell</b>	<b>55</b>
7.1 Inntektselastisitet . . . . .	55
7.2 Tolkning av inntektselastisiteten . . . . .	62
7.3 Robusthet og utvidelser . . . . .	63
<b>8 Oppsummering og konklusjon</b>	<b>67</b>
<b>Referanser</b>	<b>70</b>
<b>A Appendiks</b>	<b>75</b>
A.1 Bilde av spilleregler på panteautomat . . . . .	75
A.2 Korrelasjonsmatrise . . . . .	76
A.3 Variasjon mellom og innad i tverrsnittenhetene . . . . .	76
A.4 Komplette Tabeller . . . . .	77
A.5 Heterogenitet over tid . . . . .	80
A.6 Robusthet MHT alternativ inntektsmål . . . . .	80

**Tabeller**

1	Suits-indekser . . . . .	35
2	Suits-indekser med ulike inntektsmål . . . . .	36
3	Test av restriksjon . . . . .	52
4	Pooled OLS . . . . .	56
5	Random Effects . . . . .	58
6	Fixed Effects . . . . .	60
7	Random Effects med Interaksjonsledd . . . . .	64
8	OLS på Tverrsnitt . . . . .	65
9	Ulike inntektsmål . . . . .	65
10	Gjennomsnittlig inntekt, Random Effects . . . . .	66
11	Korrelasjonsmatrise . . . . .	76
12	Between- og Within-variasjon . . . . .	76
13	Pooled OLS, Komplette tabell . . . . .	77
14	Random Effects, Komplette tabell . . . . .	78
15	Fixed Effects, Komplette tabell . . . . .	79
16	Pooled OLS med Interaksjon . . . . .	80
17	Gjennomsnittlig inntekt, Fixed Effects . . . . .	80
18	Gjennomsnittlig inntekt, Pooled OLS . . . . .	80



# 1 Introduksjon og motivasjon

Pengespill er utbredt i Norge, og utgjør en betydelig utgiftspost for store deler av befolkningen. Lotteri og stiftelsestilsynet (2016) opplyser om at 2,2 millioner nordmenn spilte pengespill for 40,7 milliarder kroner i 2016. Dette utgjør en gjennomsnittlig innsats på 9900 kroner per innbygger over 18 år. 30,6 milliarder kroner blir tilbakebetalt som premier, noe som gir oss et gjennomsnittlig økonomisk tap på 2500 kroner per myndige innbygger. Alle pengespill i Norge er strengt regulert, med kun et fåtall av tilbydere. Kulturdepartementet har hovedansvaret for pengespill i Norge, og forvalter alle pengespill gjennom Lotteri- og stiftelsestilsynet. Kulturdepartementet forvalter også statens eierskap av Norsk Tipping som har enerett på alle de største pengespillene i Norge.<sup>1</sup> Pantelotteriet har siden 2008 vært en av svært få store nasjonale lotterier, og har siden lanseringen vært eneste privatdrevne landsdekkende lotteri med konsesjon etter §7 i lotteriforeskrifen. Kulturdepartementet vedtok i februar 2015 å utlyse fem konsesjoner for landsomfattende lotterier drevet av samfunnsnyttige og humanitære organisasjoner. Konsesjonene har en varighet på ni år, og gir tillatelse for en årlig omsetning for inntil NOK 300 millioner (Jf. Forskrift til lov om lotterier m.v. 2015 §7).

Med fem landsomfattende lotterier drevet av samfunnsnyttige og humanitære organisasjoner, vil det være mulig med et landsdekkende spillmarked for store lotterier på NOK 1,5 milliarder. Dette spillmarkedet kan ikke sees som en direkte konkurrent til de mer tradisjonelle lotteriene drevet av Norsk Tipping, da forskriftene for store lotterier drevet av ideelle organisasjoner setter andre krav og rammer for drift av lotteri og hvilke formål lotteriets overskudd skal gå til. §7 i Forskrift til lov om lotterier m.v. setter blant annet at *lotteriet kan ikke ventes å medføre problem i form av spilleavhengighet* som et vilkår for tildeling av lotteriløve. Både kravet om drift av ideelle organisasjoner, toppgevinstgrense på NOK 2 millioner og pengespill uten fare for spilleavhengighet gjør at forskning gjort på lotterier drevet av en statlig enetilbyder ikke lenger er like relevant og anvendelig. Skal vi forstå hvordan et slikt spillmarked vil fungere og hvilke konsekvenser det har for spillerne, spilltilbyderne og samfunnet, må vi ta i bruk empiri og erfaringer fra pengespill med liknende vilkår og rammer.

---

<sup>1</sup>Bingo, pokerturneringer og andre mindre pengespill drives ikke av Norsk Tipping, men reguleres av Lotteri- og stiftelsestilsynet.

## 1.1 Pantelotteriet

Pantelotteriet ble først lansert i 2008 med en utrulling i alle landets Rema 1000-butikker, og har siden blitt introdusert i nær 2400 panteautomater i flere av landets største dagligvarekjeder. Ved pant av flasker og bokser vil man ved en panteautomat med pantelotterifunksjonen kunne velge den ordinære flaskepanten på NOK 1 per pantbare beholder under 0,5 liter og NOK 2,5 for beholdere over 0,5 liter<sup>2</sup>, eller å kjøpe lodd for pantepengene. Hvert lodd koster 50 øre og kan kun kjøpes på panteautomaten for penger man får ved retur av tomgods. Hvert lodd gir spilleren muligheten til å vinne premier på en verdi mellom NOK 50 og NOK 1 000 000. Utbetalingsandelen til lotteriet er på 35%, og overskuddet går til Røde Kors. Overføringene til Røde Kors deles mellom det nasjonale Røde Kors og lokallaget hvor panteautomaten befinner seg. Alle spillere blir opplyst om lotteriets regler, utbetalingsandel og formål gjennom tydelig merking av panteautomatene.<sup>3</sup>

Pantelotteriets årlige omsetning har siden lanseringen økt fra NOK 25 907 754 i 2010 til NOK 100 412 436 i 2016. Dette skyldes at pantelotteriet har blitt mer tilgjengelig med en økning i antall panteautomater med spillfunksjon fra 730 automater i januar 2010 til 2137 i desember 2017. I tillegg til økt tilgjengelighet har andelen av pantepenger som blir brukt på lotteriet økt jevnt fra 5,74% i 2010 til 10,09% i 2016. Overføringene fra pantelotteriet har blitt en betydelig inntektskilde for Røde Kors, som har fått overført NOK 170 millioner siden 2008<sup>4</sup>, samtidig som det raskt har blitt en nasjonal lotteritilbyder, med 7 920 227 spillsesjoner i 2016.

## 1.2 Lotteri som beskatning

Som med alle statlig styrte aktiviteter, er det av interesse å se på velstand- og kostnadsfordeling av pengespill. Siden kjøp av pengespill er frivillig<sup>5</sup> kan det virke urimelig å se på det forventede tapet som en beskatning, og at begreper som skattebyrde og skatteinsidens ikke er anvendbare når vi betrakter hvem som kjøper og taper penger på lottrier. Clotfelter & Cook (1987, s. 6) argumenter for at lotterier bør ansees som en beskatning av lotteriets brukere da et stort antall nasjonale lotterier er statseide og drives med den hensikt å finansiere eller

---

<sup>2</sup>Pantesatsene vil oppjusteres til NOK 2 for beholdere under 0,5 liter og NOK for beholder over 0,5 liter, september 2018.

<sup>3</sup>Se Figur 13 i appendikset for fotografi av informasjonsplakat som er klistret på alle panteautomater

<sup>4</sup>Tall fra Olav Thon Gruppen sin årsrapport fra 2016

<sup>5</sup> Patologisk pengespillavhengighet har vært en anerkjent psykiatrisk lidelse siden den ble inkludert i den psykiatriske diagnosemanualen DSMIII i 1980. (Palleen et al. 2014, s. 12)

subsidiere fellesgoder eller veldedige formål. Videre argumenterer de for at staten bestemmer skattesatsen ved å regulere gevinstandelen. Dette gjør at pengespillenes forventede tap bør betraktes som en skatt eller avgift på lik linje med andre overføringer fra borger til stat.

Pantelotteriet og de andre store lotteriene som har fått konsesjon skal riktig nok bli drevet av samfunnsnyttige og humanitære organisasjoner, og ikke av statseide Norsk Tipping. Med statlig bestemte retningslinjer for drift og krav om at overskuddet kun skal benyttes til å finansiere samfunnsnyttige og humanitære formål, vil det med Clotfelter & Cook (1987, s.6) sin argumentasjon være hensiktsmessig å betrakte lotterienes forventede tap som en implisitt skattesats for lotterispill. På lik linje med andre skatter og avgifter er det av interesse å se hvilke inntektsgrupper som påtar seg denne skattebyrden. Dette gjøres ofte ved å identifisere om skatten er *progressiv* eller *regressiv*, altså om skatten i forhold til inntektstørrelsen er høyest hos høyinntektsgrupper eller lavinntektsgrupper.

### 1.3 Oppgavens formål

Hovedformålet for denne oppgaven er å analysere fordelingsvirkninger av lotterier og pengespill drevet av samfunnsnyttige og humanitære organisasjoner. Jeg vil undersøke om lotterier med relativt lav premieandel<sup>6</sup> og krav om å ikke forårsake spillavhengighet skiller seg fra tradisjonelle statsopererte lotterier, eller om de føyer seg inn i rekken av lotterier med regressiv skatteinsidens og en lav inntektselastisitet. Med bruk av data fra det første lotteriet av denne typen, pantelotteriet, vil denne oppgaven bestå av kvantitative undersøkelser av pantelotteriets fordelingseffekter, inntektselastisitet og beskatning.

### 1.4 Disposisjon

Denne oppgaven er strukturert i 8 kapitler, hvor vi i Kapittel 2 går gjennom tidligere forskning og teoretiske rammeverk for konsum av pengespill og lotteriers fordelingseffekter. I Kapittel 3 vil det bli presentert to statistiske metoder som blir hyppig brukt ved mål av skatteinsidens og inntektselastisitet. Kapittel 4 vil presentere denne oppgavens datamateriale og en diskusjon av hvilke kilder og forklaringsvariable det er mest hensiktsmessig å inkludere. Dette kapitlet vil også presentere resultatene fra pantelotteriets Suits-indeks, som er et deskriptivt mål på fordeling av skattebyrde. Videre vil kapittel 5 presentere det økonometriske ramme-

---

<sup>6</sup>Jf. Forskrift til lov om lotterier mv. §7 må gevinstene utgjøre minst 25 % av omsetning, i pantelotteriet er premieandelen 35%

verket den statistiske analysen vil være bygget på, samt hvilke økonometriske utfordringer som er heftet ved denne oppgavens datasett. Kapittel 6 vil presentere empiriske modellspekifikasjoner mens Kapittel 7 vil presentere de empiriske resultatene. Dette vil bli fulgt av en kort diskusjon og konklusjon i Kapittel 8.

## 2 Teorigrunnlag og litteraturgjennomgang

Pengespill og lotteri er et økonomisk felt som har blitt studert helt siden fagets begynnelse. Daniel Bernoulli (1700-1782) viste i 1738 at man kunne konstruere pengespill med forventet positiv gevinst som samtidig ville være lite attraktive å spille. Den rådende økonomiske teorien tilsa at rasjonelle aktører ville delta i alle pengespill med forventet positiv gevinst. Bernoulli (1954) forklarte denne "urasjonelle" oppførselen ved at vi ikke betrakter den forventede gevinsten, men den forventede *nytt* når vi vurderer et pengespill. Med dette introduserte Bernoulli nytte som en sentral økonomisk størrelse og postulerte at grensenytten var fallende med størrelsen på gevinsten. Siden har pengespill og andre valg under usikkerhet vært en sentral del av den økonomiske litteraturen, og har blitt brukt som teoretiske rammeverk for å forklare mennesker og andre økonomiske aktørers valg, preferanser og strategier. Lotterispill blir i denne sammenheng ofte behandlet som et naturlig eksperiment for å teste disse økonomiske teoriene.

### 2.1 Hvorfor deltar vi i lotterier?

Som nevnt i innledningen er lotteri og pengespill svært utbredt i Norge, med en total omsetning på 40,7 milliarder fordelt på 2,2 millioner nordmenn. Hvorfor vi tilsynelatende sløser bort penger på spill med forventet tap er et spørsmål som har utfordret tenkere helt siden Bernoulli. I moderne forskningslitteratur blir konsum av lotteier og pengespill blir gjerne forklart med to teoretiske innfallvinkler, lotteri som et *investeringsobjekt* og som et *konsumgode*.

#### 2.1.1 Lotterispill for økonomisk gevinst

Clotfelter & Cook (1990, s.109) viser til at lodd ofte er de billigste og lettest tilgjengelige investeringsmulighetene store deler av en befolkning har. Dette gjør at selv lotterier som har en gevinstandel på 50% av omsetning, representerer eneste mulighet til å drastisk forbedre den økonomiske situasjon for svært mange. Lodd kan dermed betraktes som et investeringsobjekt med en *liten mulighet for stor avkastning, og en stor mulighet for et lite tap* (Markowitz



1952). Med en forventet negativ avkastning på rundt 50% av innskudd <sup>7</sup> er det åpenbart at lotterispill for de aller fleste er en investering man helst bør unngå.

### 2.1.2 Risikoaversjon og risikosøkere

Holdninger til økonomisk risiko blir i litteraturen gjerne illustrert ved en enkel modell for valg av sikre og usikre økonomiske utfall (Varian 2010, s.222).

Her betrakter vi en beslutningstaker som har nyttefunksjonen  $U(W)$ , hvor formue,  $W$ , er eneste argument. Nyttefunksjonen antas å være kontinuerlig med kontinuerlige første- og andre-ordens partielle deriverte. Et individs risikoholdning kan utledes ved å se hvordan beslutningstakeren betrakter den forventede nytten,  $E(U(W))$  til et usikkert økonomisk utfall i forhold til nytten av forventningsverdien til det økonomiske utfallet,  $U(E(W))$ . Dette kan eksemplifiseres ved å betrakte et lotteri som et slikt økonomiske valg. Lotteriet har to utfall; tap, hvor beslutningstakeren vil ende på velferdsnivået  $W_1$ , eller gevinst, som fører til et høyere velferdsnivå,  $W_2$ . Sannsynligheten for gevinst er gitt som  $P$ , mens sannsynligheten for tap er gitt som  $(1 - P)$ , hvor  $0 < P < 1$ . Vi antar at lotteriet er *rettferdig*, altså at den forventede gevinsten er like stor som innsatsen.

Dette gjør at vi kan skrive lotteriets forventningsverdi,  $E(W)$ , som:

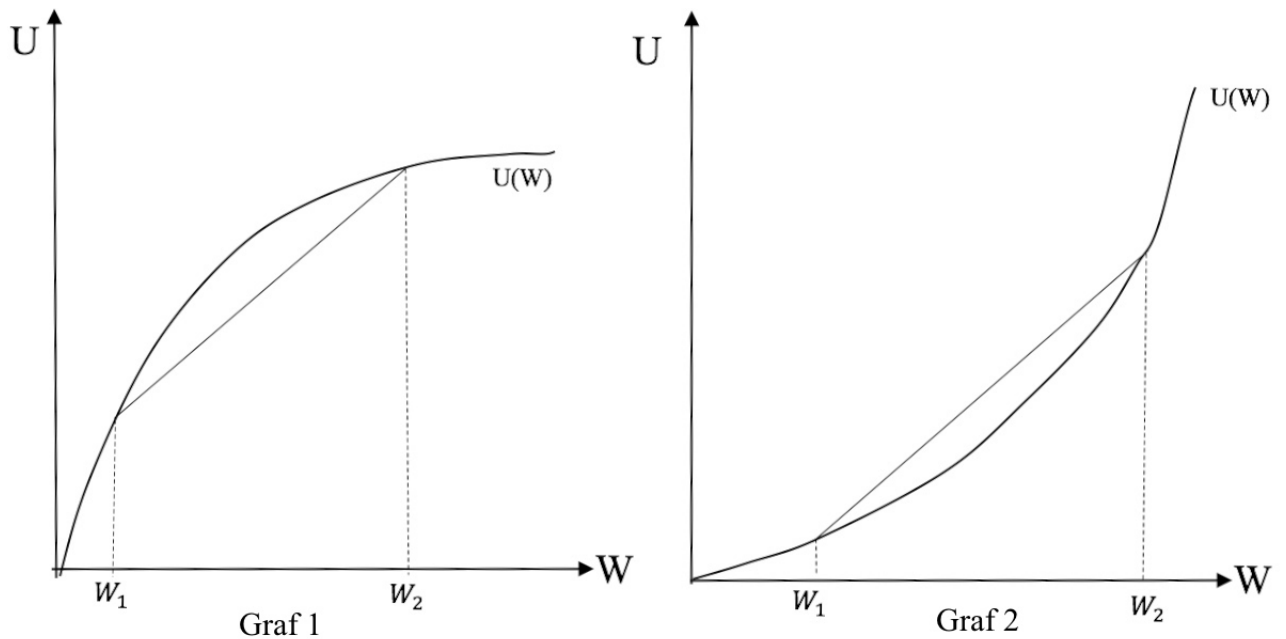
$$E(W) = (1 - P)W_1 + PW_2 \quad (1)$$

Antatt at beslutningstakeren er nyttemaksimerende, vil hun alltid foreta det valget som gir høyest nytte. Om hun velger et sikkert utfall ved å ikke delta i lotteriet eller det usikre binære utfallet til lotteriet, avhenger av om nyttefunksjonen,  $U(W)$ , er avtagende eller tiltagende med størrelsen på formue,  $W$ .

Dersom  $U(W)$  er positiv og avtagende,  $U'(W) > 0$ ,  $U''(W) < 0$ , vil nyttefunksjonen anta den konkave formen vi kan se i Graf 1, Figur 1.

Denne formen på nyttefunksjonen gjør at nytten til lotteriets forventningsverdi alltid vil være høyere enn den forventede nytten av lotteriet,  $U(E(W)) > E(U(W))$ . Vi kan se dette ved at nyttekurven i Graf 1 ligger høyere enn lotteriets forventede nytte for alle nivåer av  $P$  i intervallet  $[0 < P < 1]$ .

<sup>7</sup>Lotto-spillet til Norsk Tipping har en tilbakebetalingsprosent på 50% (Norsk Tipping 2013). Dette er standard premieandel for statlig drevne lotterier (Clotfelter & Cook 1990).



Figur 1: Graf 1 viser et risikoavert individ med en konkav nyttefunksjon, mens Graf 2 viser et risikosøkende individ med konveks nyttefunksjon.

$$U((1 - P)W_1 + PW_2) > (1 - P)U(W_1) + PU(W_2) \quad (2)$$

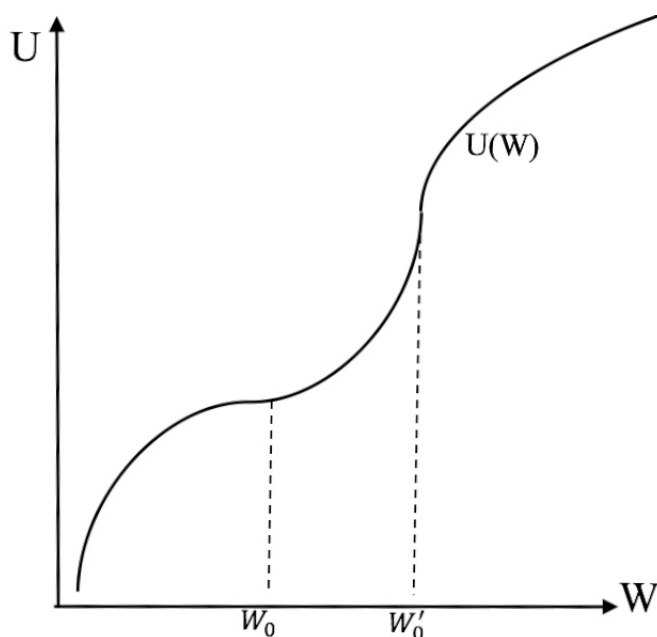
Når nyttefunksjonen er konkav mhp.  $W$ , vil beslutningstageren alltid foretrekke et sikkert utfall fremfor et usikkert utfall, selv om begge disse utfallene har lik forventningsverdi. Likning (2) viser slik risikovurdering. Dette kalles gjerne for *risikoaverse* individer i litteraturen.

Motsatt, ved en positiv og tiltagende nyttefunksjon,  $U'(W) > 0$ ,  $U''(W) > 0$ , vil nyttefunksjonen være konveks mhp.  $W$ . For dette *risikosøkende* individet, vil nytten til lotteriets forventningsverdi alltid vil være lavere enn den forventede nytten av lotteriet,  $U(E(W)) < E(U(W))$ . Dette vises i Graf 2, hvor den forventede nytten til lotteriet er høyere enn nyttekurven for alle nivåer av  $P$ .

$$U((1 - P)W_1 + PW_2) < (1 - P)U(W_1) + PU(W_2) \quad (3)$$

Et individ med en nyttefunksjon som er konveks MHP.  $W$ , vil dermed alltid velge et usikkert økonomiske utfall fremfor et sikkert utfall, så lenge disse utfallene har lik forventningsverdi (Varian 2010, s.227).

I den økonomiske litteraturen blir nyttefunksjonen som regel antatt å være konkav med velferdsnivået, altså at vi som beslutningstagere stort sett er risikoaverse. (Levy 2015, s.323). I

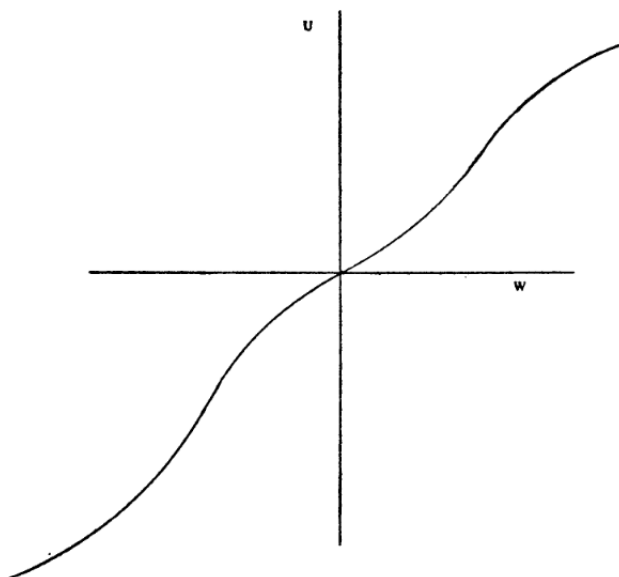


Figur 2: Friedman-Savage nyttekurve (Friedman & Savage 1948)

mange sammenhenger er dette en nyttig antagelse. Men antagelsen om strengt risikoaverse individer er lite forenelig med en virkelighet hvor vi stadig observerer risikosøkende adferd, som for eksempel spill på lotterier med forventet negativ avkastning.

En hypotese som forklarer tilsynelatende irrasjonelle kjøp av lodd med forventet tap ble fremlagt av Friedman & Savage i 1948. Her argumenterer de for at nyttefunksjonen ved enkelte mellominntektsnivåer er positiv og konveks med størrelse på formue, altså at folk i dette inntektsintervallet er risikosøkende. En Friedman-Savage nyttefunksjon vil dermed rasjonalisere kjøp av investeringobjekter med negativ forventet avkastning fordi den forventede endringen i nytte vil være positiv. Begrunnelsen bak denne ukonvensjonelle nyttekurven hevder de er at forskjellige inntektsnivåer er knyttet til ulike sosioøkonomiske samfunnslag. Ved enkelte inntektsnivåer vil man da stå på grensen mellom to samfunnslag, og små inntektsforskjeller avgjør hvilket sosialt lag man assosieres med. I dette inntektsintervallet vil marginalnyttene av formue være høy og tiltagende fordi den assosieres med et hopp eller fall mellom to samfunnssjikt. I Figur 2 er dette illustrert med en nyttefunksjon som er konkav for velferdsnivåer under  $W_0$  og over  $W'_0$ , mens den er konveks for velferdsnivåer mellom  $W_0$  og  $W'_0$ .

Ved å se på lotterispill med forskjellige størrelser på toppgevinsten, viser Clotfelter & Cook (1987, s.538) at pengespill med toppgevinster i millionsjiktet er mindre regressive enn pengespill med lave toppgevinster. Altså er konsumenter fra høyere inntektsgrupper mind-



Figur 3: Markowitz nyttekurve med referanseinntekt (Markowitz 1952)

re representert i pengespill med lavere gevinster. Dette underbygger hypotesen til Friedman og Savage om at folk kjøper pengespill hvor toppgevinsten er stor nok til å betraktelig bedre ens økonomiske situasjon, mens man avstår fra spill som ikke representerer muligheten for en betydelig endring i materiell velstand. Borge (2003, s.7) finner liknende effekter i sin studie av det norske spillmarkedet, hvor spillene med høyest konsentrasjon av kunder fra lavinntektsgrupper er spill med lav innsats og lave premier, som skrapelodd.

Friedman-Savage nyttekurven har vært gjenstand for kritikk helt siden den ble lansert i 1948. En av Friedmans studenter, Harry Markowitz, argumenterte kun få år senere at en risikosøkende mellominntektsgruppe ikke bør forklares gjennom ønsket om et hopp i sosiale klasser, og at et individs nyttenivå ikke bør måles i absoluttverdier, men som avvik i inntekt relativt til en referanseinntekt. Markowitz (1952, s.154) lanserte hypotesen om at en nyttefunksjon er konveks for inntekter under et referansepunkt og konkav over denne referanseinntekten. Dette er vist i Figur 3. Konveksiteten i nyttefunksjonen bruker han så til å rasjonalisere hvordan et nyttemaksimerende individ kan opptre risikosøkende i enkelte situasjoner og risikoavert i andre. Denne hypotesen er det teoretiske grunnlaget Kahneman & Tversky (1979) tok i bruk da de utviklet det adferdskonomiske fagfeltet *Prospect Theory*.

Det finnes også enkelte tilfeller hvor spill på lotterier får en forventet positiv avkastning. Dersom toppgevinsten ikke blir vunnet, blir den i mange lotterier overført til neste trekning. Ved gjentatte trekninger uten en vinner, kan topppremien bli større enn det totale loddsalget

for enkelte trekninger, og et lodd vil dermed få en forventet positiv avkastning. En slik situasjon oppsto i det britiske nasjonalotteriet januar 2016 etter toppremien hadde blitt overført til neste trekning 14 ganger, og nådd en størrelse på £58 millioner. For et lodd til £2, som vanligvis hadde en forventningsverdi på 95 cent, kunne lotterispillerne nå forvente en utbetaling på £3,2<sup>8</sup>. Med en forventet fortjeneste av pengespill ser vi at det kan oppstå situasjoner hvor det lodd kan betraktes som en rasjonell økonomisk investering, selv for risikoaverse individer. Oster (2004, s.8) finner i en studie av størrelse på toppgevinst og skatteinsidens at en økning i forventet avkastning som følge av gjentatte overføringer av toppgevinsten reduserer regressiviteten av lotteriets beskatning. Oster foreslår at høyere inntektsgrupper er mer sensitive for endringer i forventet avkastning som en forklaring på dette.

Alt i alt tilhører de fleste forsøk på å rasjonalisere lodd som et investeringsobjekt enten spesialtilfeller hvor den forventede avkastningen ytterst få ganger blir positiv, eller ukonvensjonelle økonomiske tolkninger om hvordan nytten blir bestemt av inntekt. Dette gjør at det i større grad vies fokus til lotterispill som et konsumgode fremfor investeringsobjekt.

### 2.1.3 Lotteri som konsumgode

Deltagelse i pengespill og lotterier med forventede tap forklares gjerne gjennom å se på nytten av selve spillhandlingen fremfor den forventede nytten av spillutfallet. Pengespill er ofte designet nettopp for å underholde og gi spillglede, også til de som taper. Lotterier er et godt eksempel på dette. Med trekninger lagt til enkelte ukedager, gir kjøp av et lodd en legitim rett til å fantasere om hvordan man skal nyte gevinsten frem til lotteriet blir avgjort. Tv-sendte trekninger i beste sendetid gjør også selve avgjørelsen til en sosial begivenhet. Mange økonomer deler dermed nytten av et lotteri inn i to komponenter; en sikker komponent som er gleden ved å delta i spillet og en monetær komponent som er den forventede premien.

Conlisk (1993) viser i sin modellering av nytte og gambling at et risikoavert individ vil kjøpe lodd dersom gleden av å spille overgår det forventede tapet av innsatsen. Conlisk mener at forholdet mellom nytte og inntekt er tilnærmet lineært over små intervaller, og at et tilsynelatende risikoavert individ vil ta risikonøytrale valg ved pengespill med en liten innsats, som lotteri. Ved risikonøytralitet trenger man kun en liten merverdi gjennom spillglede eller sosial hygge rundt spillsituasjonen for at et nyttemaksimerende individ vil delta. Problemet med en slik modellering er at lotterier stort sett har et forventet tap, og at selv risi-

---

<sup>8</sup>Estimert av Josh White, assisterende direktør i KPMGs økonomiavdeling for avisa The guardian.

konøytrale personer må kompenseres med en merverdi gjennom spillglede som overgår det forventede tapet om de skal delta i lotteriet. Kocher et al. (2014) finner gjennom eksperimenter med forskjellige lotteri-oppsett at spenning og forventning knyttet til utfallet av et lodd er utslagsgivende for motivasjonen for å delta i lotterier. Dette blir begrunnet med at en signifikant andel av studiedeltakerne selv velger å forlenge ventetiden før den eventuelle gevinsten blir avslørt, noe som indikerer at ventetiden før utfallet av lotteriet blir avslørt har en egenverdi. Samtidig velger et klart flertall av lotterideltagerne å spre de utdelte spillemidlene over flere treknings fremfor å bruke alle loddene på en gang. Dette underbygger også Conlisk sin teori om at vi tar risikonøytrale valg om innsatsen er tilstrekkelig liten.

### **Altruistisk motivert spilling**

Nasjonale lotterier har historisk sett vært en statlig inntektskilde, hvor overskuddet har blitt brukt til å dekke budsjettunderskudd eller finansiering av konkrete offentlige goder. Storbritannias første lotteri ble opprettet i 1566 for å finansiere krig mot Frankrike, operahuset i Sidney ble delfinansiert av lotterier, det samme ble universitetene Harvard og Yale (Walker 1998, s.2). I Norge har den statlige monopolisten Norsk Tipping gitt et uavkortet overskudd til samfunnsnyttige formål siden oppstarten i 1948. Overskuddet går i hovedsak til å finansiere idrett (64%), kultur (18%), og humanitære formål (18%)<sup>9</sup>. Pantelotteriets overskudd blir fordelt mellom Røde Kors nasjonalt og Røde Kors sitt lokallag i kommunen hvor panteautomaten er plassert. Felles for pantelotteriet og de andre landsdekkende pengespillene som har fått konsesjon fra kulturdepartementet er at de må bli drevet av samfunnsnyttige eller humanitære organisasjoner. Vi må derfor ikke ignorere at selv tapte penger kommer samfunnet til nytte, og at mye av motivasjonen til å delta i disse lotteriene kan være at man bidrar til en god sak. Dette blir støttet av forskning på både altruisme og lotterier som instrument for finansiering av veldedige formål. Andreoni (1990) innførte begreper som *uren altruisme* og *warm glow* i sin teoretiske modellering av private bidrag til finansiering av offentlige goder. Disse begrepene henviser at individer ikke er likegyldige til hvem som finansierer et veldedig formål, men får en større nytte ved å vite at andres velferd er et resultat av egne økonomiske donasjoner. Andreoni (1990) viser at ved å ilegge selve giverhandlingen egenverdi, kan man bedre forene økonomiske teorier og empiriske observasjoner av hvordan private donasjoner utgjør finansieringen av offentlige goder og veldedige formål. Vi må derfor ta høyde for

<sup>9</sup>Tall fra Norsk Tipping. Før overskuddet fordeles mellom idrett, kultur og humanitære formål trekkes det ut 6,4% av overskuddet til helse- og rehabiliteringsformål og 5% til spillernes eget valgte formål, grasrotandelen.

at pantelotteriets kunder i tillegg til høye gevinster og spenningen av å gamble, også motiveres av at eventuelle individuelle tap finansierer humanitære og samfunnsnyttige formål. Wiepking & Bekkers (2012, s.20) finner i en gjennomgang av empiriske studier på inntekt og veldedige donasjoner at størrelsen på donasjoner øker med inntekten. De finner også at inntektselastisiteten er mindre enn 1, altså at donasjonene som andel av den totale inntekten faller med inntektens størrelse. Dersom pantelotteriet blir betraktet som et veldedig formål fremfor et lotteri, vil dette også påvirke hvilke inntektsgrupper som kjøper lodd.

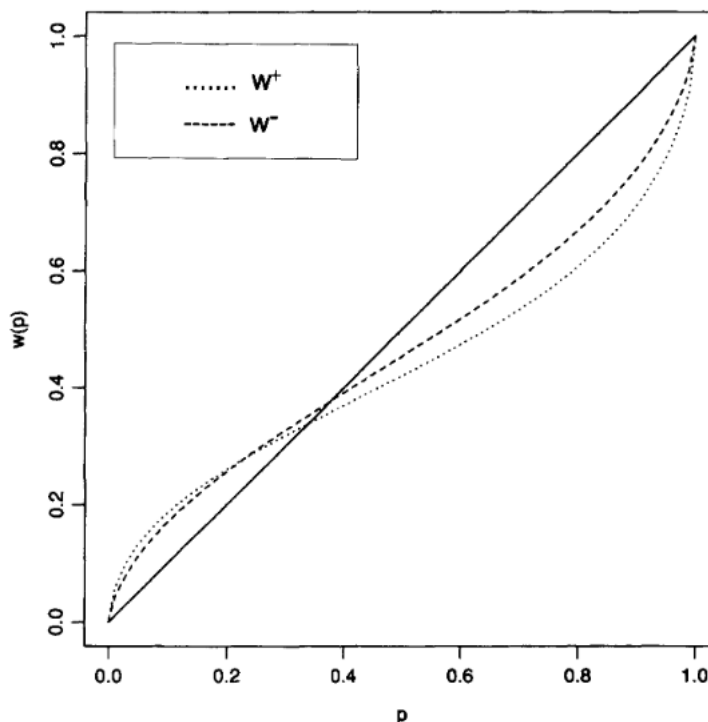
### **Adferdsøkonomiske vinklinger**

Den tilsynelatende irrasjonelle handlingen ved kjøp av lodd med negativ forventet avkastning har lenge utfordret økonomiske teoretikere. Med oppblomstringen av adferdsøkonomien har flere alternative forklaringsmodeller for risikofylt adferd blitt lansert. Adferdsøkonomien forener økonomisk teori med kunnskap fra psykologi, og tar utgangspunkt i kognitive skjevurderinger og heuristikker for å modifisere økonomiens standardantagelser om rasjonelle og nyttemaksimerende aktører. Tversky & Kahneman (1992) viser at vår subjektive forståelse av sannsynlighet avviker fra objektive sannsynligheten. Dette kommer spesielt til uttrykk i tilfeller hvor sannsynligheten for at et scenario vil inntreffe enten er veldig høy eller veldig lav. Avviket mellom opplevd sannsynlighet og faktisk sannsynlighet er illustrert i Figur 4. Tversky og Kahneman viser at vi har en tendens til å oppleve sannsynligheten for at svært sjeldne utfall, som haiangrep<sup>10</sup> eller flykrasj, som høyere enn hva sannsynligheten i realiteten er. Dette mener de også er tilfellet for vurderingen av sannsynligheten for å vinne toppremien i et lotteri. Clotfelter & Cook (1990, s.110) trekker frem at ved sannsynligheter på under en til en million, vet vi at sannsynligheten for å vinne toppremien er svært lav, men at vi har liten intuisjon på hvor små de reelle vinnerejansene faktisk er. Disse kognitive misoppfatningene av sannsynligheten til svært usannsynlige utfall, gjør at vi overvurder den forventede avkastningen til et lodd.

Det er ikke bare misoppfatninger av sannsynligheten for gevinst i lotterier som påvirker dømmekraften ved kjøp av lodd. Psykologer og adferdsøkonomer har vist at oppfatningen av at vi som spillere kan påvirke oddsen for å vinne også spiller en rolle i evalueringen av den forventede gevinsten til et lotteri. Psykologen Langer (1975) introduserte begrepet *illusjonen av kontroll* for å beskrive tendensen vi har til å overvurdere sannsynligheten for gevinst der-

---

<sup>10</sup>Kun 4 personer døde som følge av haiangrep i 2016 (Yearly Worldwide Shark Attack Survey (2017))



Figur 4: Tversky & Kahneman (1992, s.313) viser at en subjektivt vektet sannsynlighet,  $w(p)$ , avviker fra faktisk sannsynlighet,  $p$ . Linjene  $w^+$  og  $w^-$  viser hvordan vi vektet sannsynlighetene for hendelser med henholdsvis gevinst og tap.

som vi selv har gjort en innsats for å velge "riktig" lodd. Deltagerne i Langers studier valgte blant annet å beholde lodd eller tallrekker til lotterier de selv hadde plukket ut da de ble tilbudt lodd med statistisk større sannsynlighet for å vinne. Disse effektene har også blitt vist i reelle lotterisituasjoner hvor spillere øker sin subjektive sannsynlighetsoppfatning for vinnerjanser i butikker som nylig har solgt et vinnerlodd (Guryan & Kearney 2005, s.469).

Pantelotteriet skiller seg fra alle andre lotterier ved at betalingsmiddelet kun kan være tomme flasker eller bokser. Denne vesentlige forskjellen gjør at økonomisk teori om lotterier og pengespill ikke nødvendigvis lar seg anvende når vi betrakter pantelotteriet. Ved antagelsen om at pantepenger er perfekte substitutter for vanlig valuta, hadde vi hatt større grunn til å anse funn i litteraturen om pengespill og lotterier som gyldige for pantelotteriet. Det er derimot flere grunner til å ikke sidestille pantepenger med vanlige penger på konto. En grunn er at mange kan oppfatte det som flaut å be om småpenger for å resirkulere flasker. I det norske pantesystemet blir man regelrett tvunget til å motta penger for noe som av enkelte kan ansees som en god gjerning. Titmuss et al. (1970) argumenterer i deres bok om bloddonasjoner at en kompensasjon i form av penger reduserer antall blodgivere, da mange finner det ubehagelig å ta betalt for noe som burde være en altruistisk handling. Denne hypotesen har



også blitt foreslått for flaskepant. Gneezy et al. (2013, s.22) argumenterer i boka "The Why Axis" for at en kompensasjon for å returnere brukte flasker omdefinerer resirkuleringen av flaskene fra en god altruistisk handling til en handling gjort av egne økonomiske interesser. Videre argumenterer de for at dette vil få vedkommende som returnerer de brukte flaskene til å fremstå som gjerrig. Ved antagelsen om at resirkulering av flasker oppfattes som en altruistisk handling, kan vi se for oss at flere velger å "kvitte" seg med pantepengene gjennom pantelotteriet fremfor å kreve pengekompensasjonen i kassa. Vi bør derfor ikke betrakte tilgodelappen for pantepenger og penger på en bankkonto som perfekte substitutter.

## 2.2 Lotterier og fordelingseffekter

Staten er i mange land enetilbyder av lovlige pengespill. Med en betydelig risiko for spillavhengighet og store personlige økonomiske tap (Griffiths & Wood 1999, s.20), hviler det derfor et ansvar på myndighetene for å forstå hvem som spiller. Det har følgelig blitt gjennomført bred forskning på pengespill og dets konsekvenser for både enkeltpersoner og samfunn. På samfunnsnivå har skatteinsidens og regressiv beskatning lenge vært en sentral del av litteraturen om lotterier og pengespill. Spiro (1974) og Brinner & Clotfelter (1975) var de første til å gjøre empiriske studier av hvordan statlige lotterier beskatter ulike inntektsgrupper i befolkningen, og viste at statlig drevne lotterier er regressivt beskattende. Disse resultatene har senere blitt reproduisert i en rekke land og for et vidt spekter av pengespill (Price & Novak 2000, s.83).

Svært få studier på lotterier og fordelingseffekt identifiserer en proporsjonal eller progressiv beskatning. Enkelte studier som analyserer lotteriers skatteinsidens over tid, har funnet progressivt beskattende lotterier i enkelte år i utvalgsperioden. (Miyazaki et al. 1998, s.164). Andre studier som viser progressivitet er gjerne spesialtilfeller, som ved svært høye premier. Oster (2004, s.9) viser at dersom størrelsen på toppgevinsten blir tilstrekkelig stor, er det en teoretisk mulighet for progressivt beskattende lotterier. Spill som krever større grad av egen innsats og som premierer kunnskap, som oddsspill har også en mer proporsjonal fordelt skattebyrde (Borge 2003, s.5). Begge disse effektene gjør at pengespill hvor utfallet blir umiddelbart avklart ved kjøp, som skrapelodd eller enarmede banditter, er spesielt regressivt beskattende. Dette er også dokumentert i flere empiriske studier (Price & Novak 2000, s.89). Med et umiddelbart avklart utfall og en relativt lav toppgevinst på en NOK 1 000 000, er det nærliggende å tro at disse effektene også vil trekke pantelotteriet i en mer regressivt

beskattende retning.

Alt i alt viser nær all empirisk forskning at lotterier og pengespill i større grad blir kjøpt og spilt av lavere inntektsgrupper, noe som gir nettopp disse inntektsgruppene en disproporsjonal skattebyrde. Litteraturen på dette feltet viser klart at lotterier som statlig inntektskilde eller som overføringer til veldedige organisasjoner har en regressiv skatteinsidens.



## 3 Metode

For finne ut om pantelotteriet føyer seg inn i rekken av lotterier og pengespill som har en regressiv skatteoverveltning, vil jeg analysere lotteriets fordelingseffekter ved bruk av to statistiske metoder, Suits-indeksler og økonometriske modeller for å identifisere inntektselastisiteten.

### 3.1 Suits-Indekser

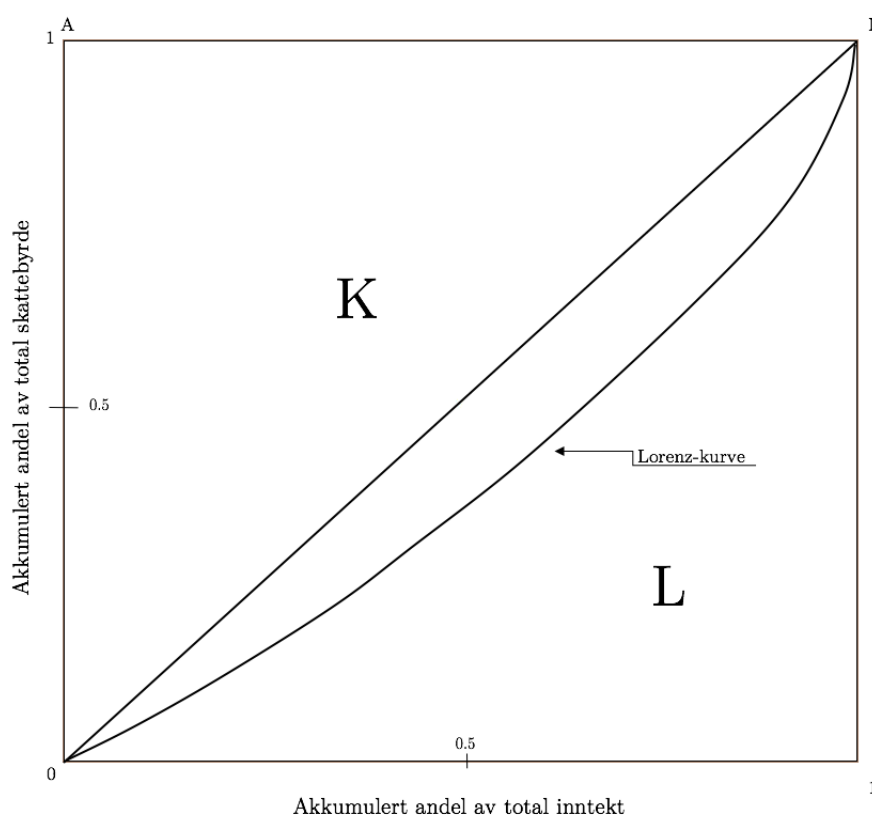
Daniel Suit konstruerte i 1977 et mål for den statistiske spredningen av skattebyrde mellom ulike innteksgrupper (Suits 1977). Bedre kjent som *Suits-indeksen*, gir dette målet oss en størrelse på fordelingseffektene til en skatt eller avgift. Med en intuitiv forståelse og en lett sammenliknbar størrelse, har Suits-indeksler blitt et populært vektøy for å beskrive fordelingseffektene av ulike skatter og avgifter. Clotfelter & Cook (1987), Price & Novak (2000), Borge (2003) og Hansen (1995) bruker alle Suits-indeksler som mål på implisitt skattebyrde. For å kunne sammenlikne de empiriske resultatene fra denne oppgaven med andre funn i den økonomiske litteraturen om pengespill, vil jeg inkludere Suits-indeksler som et mål på pantelotteriets fordelingseffekter.

Ved å ta utgangspunkt i oppsettet til den bedre kjente Gini-koeffisienten<sup>11</sup>, gir Suits-indeksen et mål på skatteinsidens ved å sammenlikne den kumulative belastningen av en skatt eller avgift med den kumulative inntekten i en befolkning som er rangert fra lavest til høyest inntekt. På samme måte som Gini-koeffisienten, lar Suits-indeksen seg fremstille grafisk på en svært intuitiv måte. I Figur 5 er utvalget rangert etter inntektsstørrelse fra lavest til høyest. Langs X-aksen vises den kumulative andelen av total inntekt mens Y-aksen viser den kumulative andelen av den totale skattebyrden. Den diagonale linjen, 0B, viser en situasjon med proporsjonal skatt, hvor alle skatter en like stor andel av inntekten. Forholdet mellom den akkumulerte andelen av total inntekt og den akkumulerte andelen av total skattebyrde blir vist i Figur 5 som en Lorenz-kurve. Fordelingseffekten av en skatt eller avgift blir målt som Lorenz-kurvens samlede avvik fra proporsjonalitetslinjen, 0B. Jo større det samlede avviket til Lorenz-kurven er fra proporsjonalitetslinjen, dess mindre proporsjonalt beskattende

---

<sup>11</sup>Gini-koeffisienten er et ofte brukt numerisk mål på den statistiske spredningen av inntekten i en befolkning. Koeffisienten blir gitt av avstanden mellom den kumulative inntekten i en befolkning rangert etter inntektsstørrelse og en situasjon hvor inntekten er uniformt fordelt. Gini-koeffisienten antar en verdi mellom 0 og 1, hvor en koeffisient lik 0 indikerer en uniformt fordelt inntekt mens 1 indikerer at all inntekt er plassert hos én person.

vil den aktuelle skatten være. Dersom Lorenz-kurven befinner seg under proporsjonalitetslinjen, vil vi ha en progressiv beskatning ettersom dette betyr at befolkningen i utvalget med lavere inntekt betaler en lavere andel av den akkumulerte skattebyrden enn sin andel av den akkumulerte inntekten. Dersom Lorenz-kurven ligger over proporsjonalitetslinjen, vil den akkumulerte andelen skattebyrden til lavinntektsgruppene være høyere enn deres akkumulerte andel av inntekt, noe som gir oss en regressiv beskatning.



Figur 5: Eksempel av en Suits-indeks med en progressiv skattebyrde

Suits-indeksen ( $S$ ) tallfestes matematisk av størrelsesforholdet mellom området over proporsjonalitetslinjen, triangelet  $OAB$ , og området under Lorenz-kurven. I Figur 5 er dette henholdsvis område  $K$  og  $L$ .

Likning (4) viser nærmere at Suits-indeksen blir gitt av forholdstallet mellom det arealet mellom Lorenz-kurven og proporsjonalitetslinjen,  $K-L$ , og arealet under proporsjonalitetslinjen,  $K$ . Dette gjør at Suits-indeksen kan anta en verdi mellom  $-1$  og  $1$ . Er  $L > K$ , vil  $S = 1 - \frac{L}{K} < 0$ , og Suits-indeksen antar en verdi mellom  $-1$  og  $0$ . Dette viser en regressivitet i skattebyrdefordelingen. Er  $L < K$ , vil  $S = 1 - \frac{L}{K} > 0$ , og vi får en Suits-indeks mellom  $0$  og  $1$ , som indikerer en progressiv beskatning. Dersom vi betrakter en proporsjonalt beskattende avgift vil vi få en Suits-indeks lik  $0$ ,  $L = K \rightarrow S = 1 - 1 = 0$ .

I ekstremtilfellene hvor et offentlig gode, f.eks. en skulpturpark, blir finansiert av utvalgets rikeste person, vil  $L = 0 \rightarrow \frac{L}{K} = 0$ , og  $S = 1$ . I et tilfelle hvor utvalgets fattigste får hele skattebyrden, vil  $L = 2K \rightarrow \frac{L}{K} = 2$ , og  $S = -1$ .

$$S = \frac{K - L}{K} = 1 - \frac{L}{K} \quad (4)$$

Likning (5) viser oss at areal L blir gitt som integralet til Lorenz-kurven. Her er  $y_i$  den kumulative andelen av den totale inntekten til individ  $i$ , mens  $T(y_i)$  er den kumulative andelen av den totale skattebyrden. Populasjonen,  $N$ , er ordnet etter inntekt i stigende rekkefølge.

$$L_x = \int_0^N T(y) dy \quad (5)$$

Område K blir gitt som arealt i triangelet 0AB:

$$K = \int_0^N \frac{1}{n} + y_{i-1} \quad (6)$$

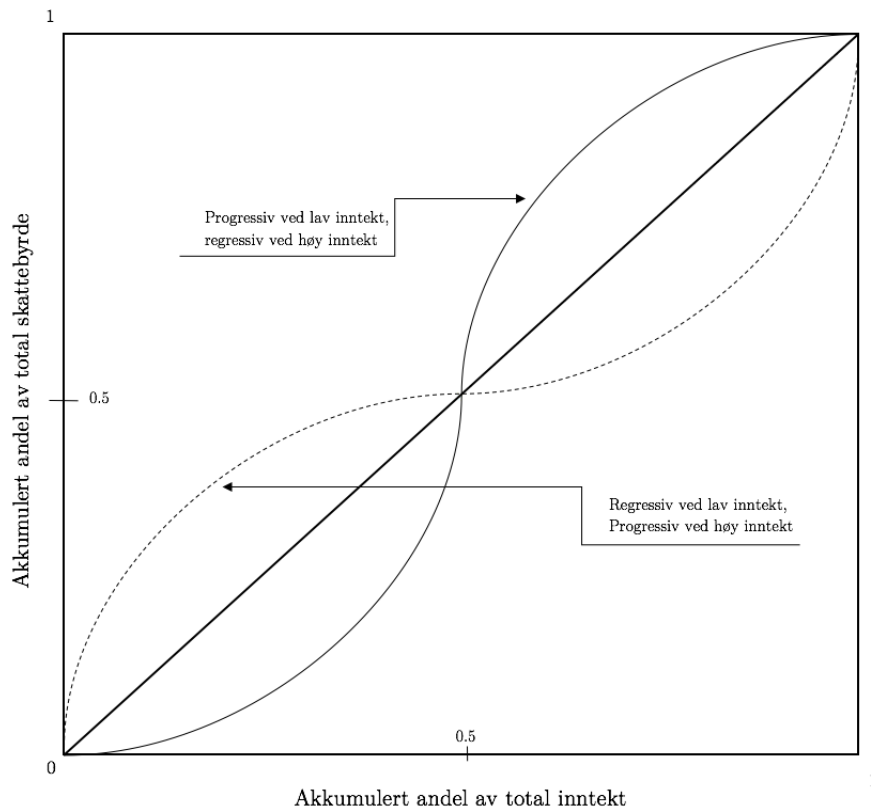
Og vil gjennom denne oppgaven bli gitt:

$$K = \frac{1}{2}(1 * 1) = 0.5 \quad (7)$$

Denne oppgavens datasett består av et diskret datamateriale. Dette gjør at vi bruker følgende tilnærming av Likning (5) for å finne størrelsen av L:

$$L = \sum_{i=1}^N (1/2)[T(y_i) + T(y_{i-1})(y_i - y_{i-1})] \quad (8)$$

Som alle mål har også Suits-indeksen klare begrensninger. Det er urimelig å kreve at et enkelt tall skal gi et nyansert bilde av noe så komplekst som fordeling av skattebyrde. Suits-indeksen viser kun gjennomsnittlig progressivitet over hele inntektsfordelingen, og vil dermed ikke hensynta at den lokale skattefordelingen kan variere mellom inntektsgruppene. Suits (1977, s.752) trekker selv frem dette problemet og illustrer det med et tilfelle hvor en skatt er regressiv ved lave inntektsnivåer og progressiv ved høye. Dersom arealene hvor Lorenz-kurven ligger under diagonallinjen er like store som arealet hvor Lorenz-kurven ligger over, vil vi få en Suits-indeks lik null, mens skattebyrden vil oppleves som skjevfordelt for store deler av befolkningen. Dette er illustrert ved den striplede kurven i Figur 6.



Figur 6: To Lorenz-kurver som begge gir  $S=0$

## Suits-indeks og pantelotteriet

Et ideelt datasett for beregning av Suits-indeks vil inneholde informasjon om hver enkelt innbyggers inntekt og betalte skatter. Dette datasettet kunne enkelt blitt bearbeidet til å finne hver enkelt innbyggers andel av befolkningens samlede inntekt og andel av samlede skatt. Videre kunne man rett frem kalkulert Suits-indeksen ved ordne populasjonen etter inntekt i stigende rekkefølge og bruke Likning (4), (7) og (8). Denne oppgavens datasett har hverken individdata på spill eller lønn, og vil dermed nyttiggjøre data på kommunenivå for å kalkulere pantelotteriets Suits-indeks.

Ved å aggregere opp solgte lodd i hver enkelt kommune, finner vi spillkonsum per capita for kommunene. Ettersom uttaksraten til pantelotteriet er konstant, vil spillkonsumet per capita fungere som mål på pantelotteriets beskatning i denne oppgaven. Videre bruker jeg brutto medianinntekt som mål på inntektsnivået i kommunen<sup>12</sup>. Ved å bruke *representative* innbyggere fra hver kommune, vil vi nå kunne ordne utvalget etter inntekt og finne tilhø-

<sup>12</sup>Dette er samme mål på inntekt som vil bli brukt til å estimere pantelotteriets inntektselastisitet senere i oppgaven, og blir grundig diskutert i Kapittel 4

rende andel av samlet inntekt og andel av samlet spillkonsum. Dette vil i sin tur anvendes til å finne akkumulert andel av samlet inntekt,  $y_i$ , og akkumulert andel av samlet spillkonsum,  $T(y_i)$ . Dette oppsettet blir anvendt av flere studier på fordelingseffekter av lotterier og pengespill. Price & Novak (2000, s.84) bruker områdedata fra den administrative enheten tilsvarende fylker for å kalkulere Suits-indekser, det samme gjør og Hansen (1995, s.390) og (Borge 2003, s.4). Hansen (1995) problematiserer bruken av områdedata og hevder at dette gir Suits-indekser nærmere null ettersom inntektsforskjeller mellom kommuner gjerne er langt mindre enn inntektsforskjeller mellom individer.

## Begrensinger

Suits-indeksen er kun et punkttestimat, og gir oss ikke muligheten til å estimere varians og konfidensintervaller eller gjennomføre hypotesetester. Dette er spesielt relevant om man ønsker å se om skatteoverveltningen endres over tid. I en utbygging av Suits statistiske rammeverk, utleder Anderson et al. (2003) en metode for å beregne variansen til Suits-indekser. Her gjennomfører de simuleringer i et nytt *pseudodatasett* ved å konstruere nye datasett gjennom gjentatte tilfeldige utvalg av det opprinnelige datasettet, såkalt *bootstrapping*. De finner at små utvalg er spesielt utsatte for å feilaktig vise en endring i skatteoverveltningen dersom det kun er noen få observasjoner med store endringer, mens store utvalg ikke er like sensitive for slike endringer. Med et datasett med inntektsinformasjonen fra over 300 kommuner, kan det argumenteres at studier av pantelotteriet ikke vil ha en overhengende risiko for å feilaktig identifisere endringer i skatteprogressiviteten over tid.

Som annen deskriptiv statistikk, begrenses også Suits-indeksen av at den ikke kan ta hensyn til at spilltilbøyeligheten også påvirkes andre faktorer enn inntekt. Dette er uproblematisk om vi kun ønsker et enkelt mål på hvordan skattebyrden fordeles mellom inntektsgrupper. Suits-indeksen gjør oss derimot ikke i stand til å isolere effekten av inntekt på spilltilbøyeligheten, da både spilltilbøyelighet og inntekt gjerne blir bestemt av andre faktorer som utdanning og alder. Det er dermed ønskelig ta i bruk statistiske estimeringsmetoder som tar høyde for dette, og som kan kontrollere inntektseffekten for effekten av andre faktorer som påvirker spilltilbøyeligheten. Ved å gjøre dette kan vi bedre finne den isolerte effekten av inntekt på spilltilbøyeligheten.



### 3.2 Inntektselastisitet

Økonometriske estimeringer som estimerer inntektselastisiteten til lotterier og pengespill er en mye brukt metode for å identifisere hvilke inntektsgrupper beskattes av lotterier. For å identifisere effekten av inntekt på spilltilbøyeligheten til brukere av panteautomater med lotterifunksjon, er det av stor nytte å utføre økonometriske regresjoner. Dette gjør at vi kan fastsette hvordan ulike sosioøkonomiske forhold, deriblant inntekt, påvirker spilladferden. Inntektselastisiteten vil vise oss den prosentvise endringen i andelen pantepenger som brukes på lotteri ved en økning i inntekten på 1%. Inntektselastisiteten vil også gjøre oss i stand til å klassifisere pantelotteriet som et normalgode eller et mindreverdige gode, altså om konsumet vokser eller faller med økt inntekt. I sin sammenlikning av inntektselastisiteten til lotterier i 82 land, finner Garrett (2001, s.222) at etterspørsel etter lodd faller med inntekt i de landene med høyest BNP per capita. Det er derfor relevant for denne oppgaven og identifisere om også pantelotteriet føyer seg inn i rekken av pengespill med negativ inntektselastisitet. Ved å følge panteautomatene i en periode over fire år, får vi et ekte paneldatasett som muliggjør modellspesifikasjoner hvor vi kan utnytte variasjonen mellom tversnittenhetene til å gi oss en mer realistisk tilnærming til de effektene vi ønsker å estimere. Vi kan også undersøke om inntektselastisiteten endres over tid ved å estimere inntektselastisiteten år for år.

Med paneldatasett kan vi nyttiggjøre oss av flere statistiske metoder for å bedre estimere presise og forventningsrette inntektselastisiteter.

## 4 Datamateriale

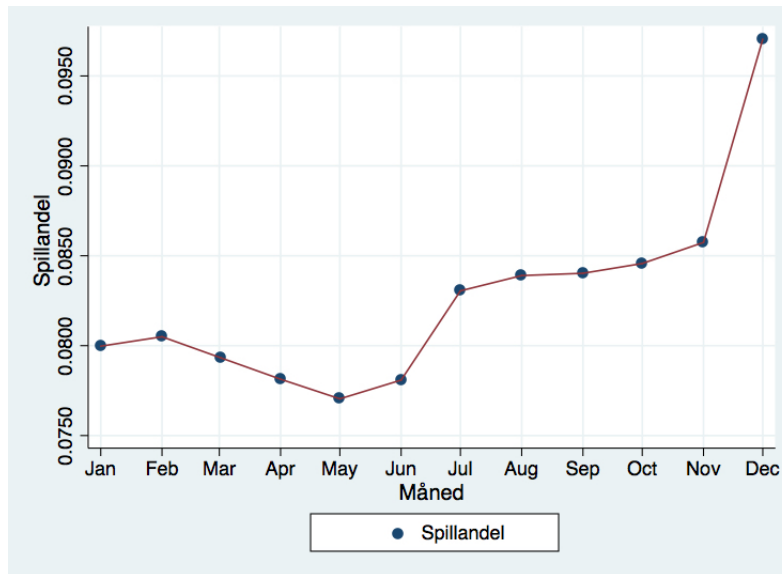
I den empiriske analysen av skatteoverveltingen til pantelotteriet tar denne oppgaven utgangspunkt i datamateriale fra pantelotteriets panteautomater, som videre blir koblet med demografiske data fra Statistisk sentralbyrå. Pantelotteriet drives av Norsk Pantelotteri AS, som eies av Røde Kors og Olav Thon Gruppen. All informasjon om pantetransaksjoner og loddsalg for alle pantelotteriets panteautomater er levert av Olav Thon Gruppen. For å analysere hvilke demografiske faktorer som er drivende for spill på pantelotteriet, kobler jeg informasjon fra pantelotteriet sammen med demografisk informasjon fra kommunene hvor panteautomatene befinner seg. På den måten får jeg konstruert et paneldatasett som er godt egnet til økonomiske analyser.

### 4.1 Data fra Pantelotteriet

Hver panteautomat registrerer samtlige pantetransaksjoner med informasjon om tidspunkt, pantebeløp, loddkjøp, og premieutbetalinger. Datamaterialet identifiserer også i hvilken butikk panteautomaten er plassert og i hvilket fylke denne butikken befinner seg. Datasettet gir ingen informasjon om de enkelte pantetransaksjonene, men gir data fra den daglige omsetningen til hver enkelt panteautomat fra 2012 til 2015. Dette lar seg videre aggregere opp til uker, måneder og år.

For å kunne se på inntektsfordelingen og andre sosioøkonomiske variable blant pantelotteriets brukere tar jeg i bruk datamateriale fra Statistisk Sentralbyrå (SSB). Datamaterialet utgis årlig, og vil dermed ikke variere fra dag til dag eller måned til måned. Av denne grunn vil jeg benytte data fra panteautomatene som er aggregert opp til informasjon for hvert enkelt år. Figur 7 viser variasjon i andelen av totalt pantebeløp som brukes på pantelotteriet over året i perioden 2012-2015. Ved å bruke årlig data, unngår vi disse sesongeffektene. Som et ganske nytt spillkonsept, er det ønskelig å se om fordelingsvirkningene endres etter hvert som pantelotteriet blir mer tilgjengelig og bedre kjent i befolkningen. Miyazaki et al. (1998, s.164) finner at lotterier blir signifikant mindre regressive over tid og over lotteriets livsløp. Det er derfor av interesse å observere pantelotteriet gjennom flere år for å se om pantelotteriet følger denne utviklingen.

For å kunne ta i bruk sosioøkonomiske forklaringsvariable må vi koble sammen informasjonen vi får fra hver enkelt panteautomat og butikk sammen med demografiske data fra de



Figur 7: Viser månedlig gjennomsnittlig andel av pantepenger som blir brukt på pantelotteriet i tidsintervallet 2012-2015, der spillandelen er definert som i Likning (9)

geografiske områdene butikkene ligger i. Man kan ta i bruk et vell av ulikt definerte geografiske områder og administrative enheter for å gjøre dette. Price & Novak (2000, s.2) tar i bruk data fra *counties*, tilsvarende fylker, når de analyserer fordelings effekter av lotterier i Texas. Samtidig viser de til studier hvor det har blitt brukt områdedata for folketellinger i Pennsylvania (Heavey 1978), postnumre i Maryland (Clotfelter 1979), og fylker i Colorado (Hansen 1995), for å måle skatteinsidens av ulike pengespill.

En mulighet er å benytte fylke som geografisk enhet. Fra datasettet levert av pantelotteriet får jeg oppgitt hvilket fylke butikkene er plassert i, noe som kun gir 19 forskjellige regioner å plassere panteautomatene i. Dette byr på flere problemer. Vi får blant annet lite variasjon i de sosioøkonomiske forklaringsvariablene vi ønsker å knytte til panteautomatene. Samtidig går vi glipp av all variasjon innad i fylkene. Price & Novak (2000, s.87) viser til at deres bruk av *counties* fremfor *municipalities*<sup>13</sup> i store grad forklarer hvorfor de får mindre regressive resultater enn sammenlignbare studier gjort på kommunenivå. Hansen (1995) peker også på samme problem ved hennes bruk av amerikanske counties. Fylker som et regionalt nivå vil med andre ord være lite hensiktsmessig dersom vi ønsker et datasett med tilstrekkelig variasjon.

Ved å bruke informasjon fra kommunenivå vil vi få et langt mer nyansert bilde av befolkningen som er knyttet til de forskjellige panteautomatene. På kommunenivå vil de demografiske variablene ha større variasjon mellom de geografiske enhetene, noe som vil gi oss en

<sup>13</sup>Counties og municipalities er amerikanske administrative enheter tilsvarende fylker og kommuner.

mer presis økonometrisk analyse. De fleste butikkjedene har oversikter over hvilke butikker som ligger i hvilke kommuner. Dette gjør arbeidet med å identifisere hvilken kommune de forskjellige butikkene befinner seg i lettere. Denne oppgavens datasett konstrueres dermed gjennom et tidkrevende arbeid, hvor samtlige av datasettets 2357 panteautomater kobles til sine tilhørende kommuner.

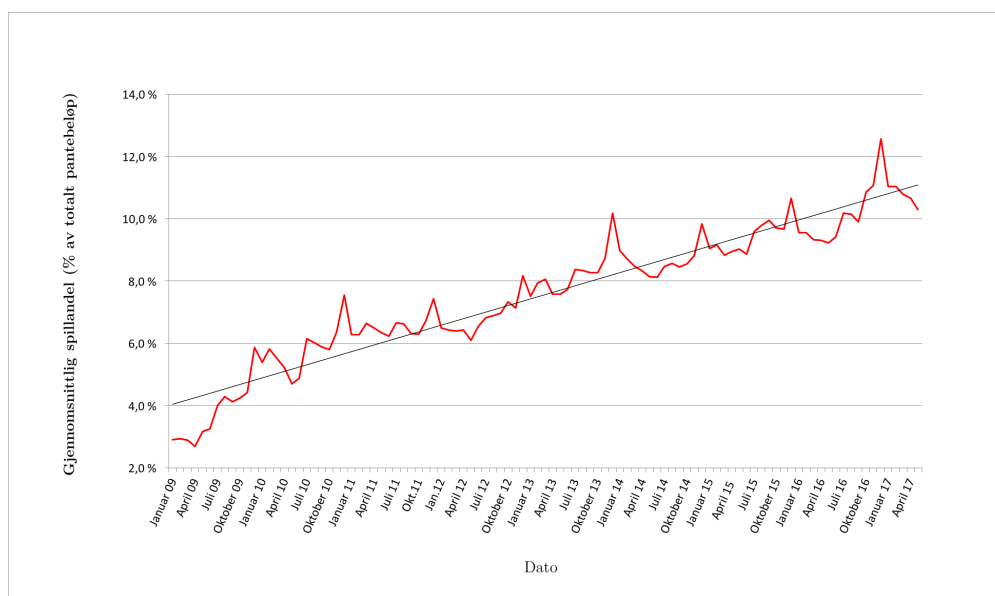
Et mulig problem med bruk av data på kommunenivå er at det vil forekomme panting på tvers av kommunale grenser. Ved å bruke kommunene til panteautomatene som geografisk enhet, vil vi få at alle pantetransaksjoner og lotterispill assosieres med demografiske data fra den respektive kommunen. Dette vil også gjelde for pantetransaksjoner som er foretatt av innbyggere fra andre kommuner. En vanlig måte å redusere dette empiriske problemet med "grenseoverskridende" adferd, er å bruke økonomiske regioner. Statistisk Sentralbyrå har delt landet inn i 89 økonomiske regioner som geografisk avgrensede områder med et felles arbeidsmarked og sammenflettet økonomi. Bruk av økonomiske regioner vil i større grad ta hensyn til at folk benytter seg av panteautomater utenfor sine hjemkommuner. Borge (2003, s.4) finner i sin undersøkelse av regressiviteten til lotterier, skrapelodd og sportspill drevet av Norsk Tipping at man får noe mer regressive resultater om man bruker økonomiske regioner fremfor kommuner. Da jeg kan finne ut hvilke kommuner de fleste av de 2357 butikkene datasettet inneholder ved å ta ibrug butikkoversikten til de forskjellige dagligvarekjedene, er dette noe jeg velger å gjøre fremfor å identifisere hvilke økonomiske regioner disse butikkene ligger i. Transportøkonomisk institutt viser i Den nasjonale reisevaneundersøkelsen (Hjorthol et al. 2014), at gjennomsnittlig lengde på innkjøpsreiser er 7.7 kilometer, mens Lavik og Jacobsen (2015, s.18) finner at 83% av dagligvarehandler blir lagt til nærmeste butikk. Det er ingen grunn til å tro at tallene for hvor man panter skiller seg i nevneverdig grad fra dette. Snarere tvert imot da dagligvarebutikker kan variere i både pris og utvalg, mens panten vil være lik i alle butikker. Vi kan derfor anta at panting på tvers av kommunegrenser ikke representerer et alvorlig problem for økonometriske analyser. Det bør noteres at det kan komme lekkasjer fra butikker som ikke har pantelotteriet, men det er ingen grunn til å tro at dette vil skje i stor grad over kommunegrensene. Price & Novak (2000, s.87) argumenterer for at bruk av fylker som geografiske enheter begrenser problemet med lekkasjer mellom kommuner ved at man nødvendigvis krysser fylkesgrenser sjeldnere enn kommunegrenser.

Etter lanseringen av pantelotteriet i 2008, har stadig flere butikker og dagligvarekjeder implementert panteautomater med lotterifunksjonen. Dette gjør at antallet tverrsnittene-

ter øker med hvert år i paneldatasettet. I 2012 er det registrert 1897 panteautomater, mens i 2015 har antall automater økt til 2357<sup>14</sup>. Ved å observere de samme automatene over opp- til fire tidsperioder vil vi kunne konstruere et paneldatasett hvor vi kan benytte oss av variasjonen over tid innad i tverrsnittene. Dette vil gi mer presise estimater, og vi kan i prinsippet ta i bruk empiriske metoder som transformerer vekk variasjon mellom tverrsnittene ved withintransformasjoner.

## 4.2 Lotterisalg (Avhengig variabel)

Med fordelingseffektene til pantelotteriet som tema for denne oppgaven, blir loddsalg den avhengige variabelen av interesse. Siden pantepenger er eneste gyldige betalingsmiddel for kjøp av lodd i pantelotteriet støter vi på enkelte problemer ved å definere loddsalget som totalt antall solgte lodd.



Figur 8: Utvikling av andel pantepenger som går til lotteriet på landsbasis i tidsintervallet 2009-2017.

Antar vi at øl, brus og annen leskedrikke er normale konsumgoder, vil vi følgelig ha at personer fra høyere inntektsgrupper vil ha flere flasker og bokser som kan pantes enn personer fra lavere inntektsgrupper. Uten informasjon om hvor stort forbruket av pantbare drikkevarer de forskjellige inntektsgruppene har, vil vi ved å betrakte totalt loddsalg risikere å overestimere effekten av inntekt, kun fordi velstående antas å ha større tilgang på pantbare flasker.

<sup>14</sup>Jeg fjerner 60 butikker fra datasettet da disse har under 100 årlige spillsesjoner. Enkelte har en kunstig høy spillandel lik 1.

Ved å bruke *andel* av totale pantepenger som blir brukt på pantelotteriet, altså andelen lodd-salg utgjør av den totale pantemengden, som avhengig variabel unngår vi dette problemet. Bruk av spillandel fremfor nominelt salg av lodd gjør også at vi unngår problemer med størrelsen til kundemassen og omsetningen til butikkene panteautomatene er plassert i. Enkelte panteautomater er plassert i butikker med langt færre kunder enn andre, noen automater er også plassert i butikker som kun er åpne på søndager. Dette gjør at variasjonen i antall solgte lodd vil i stor grad kunne forklares av butikkens kundegrunnlag, som er uobserverbar. Ved å bruke andelen av pantepenger som går til lotterispill omgår vi problemene som oppstår av variasjon i butikkens omsetning. Den avhengige variabelen blir definert som;

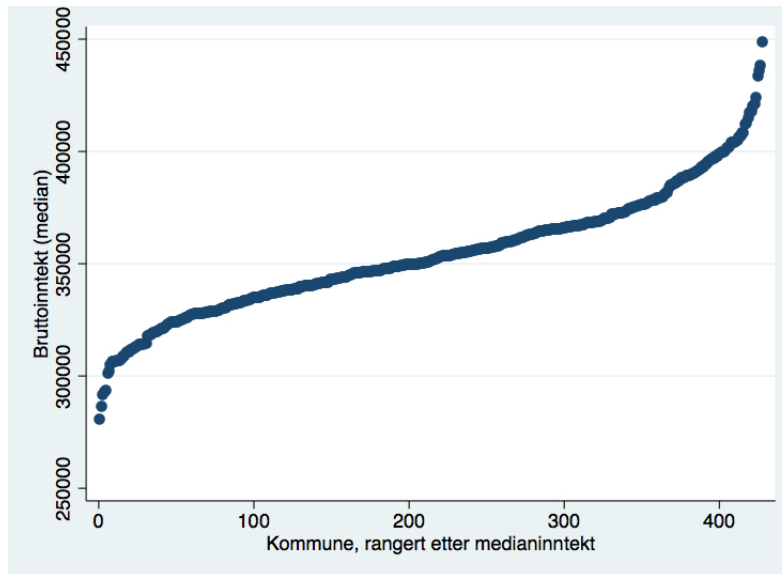
$$Spillandel = \frac{\text{NOK brukt på pantelotteriet}}{\text{Totalt pantebeløp}} \quad (9)$$

Figur 8 viser oss utviklingen i panteandelen fra lotteriets begynnelse i 2009. Vi ser en stabil positiv trend over tid med en økning gjennomsnittlig lotteriandel fra 3.73% i 2009 til 10.09% i 2016. Gjennomsnittlig panteandel for hele utvalgsperioden er 8.3%, med et standardavvik på 3.82%.

### 4.3 Inntektsvariabel

Med skattefordeling som denne oppgavens tematikk er lotterispillerens inntekt forklaringsvariabelen av størst interesse. Som vi så i litteraturgjennomgangen er nær alle lotterier og pengespill regressivt beskattende og tynger altså lavinntektsgrupper i høyere grad enn høyinntektsgrupper. For å se om dette også er tilfellet for pantelotteriet tar jeg ibruk områdedata om inntektsnivået i kommunene panteautomatene er plassert i. Jeg definerer inntekten som median bruttoinntekt i kommunen. Bruken av medianinntekt fremfor gjennomsnittsinntekt gjør at vi unngår et feilaktig bilde av butikkens kundegruppe i folkefattige kommuner med et fåtall personer med ekstremt høy inntekt. Et eksempel på dette er Frøya kommune, hvor medianinntekten i 2015 var på NOK 349 000 mens gjennomsnittsinntekten var på NOKK 482 000 (SSB,2017)<sup>15</sup>. Bruttoinntekten gir oss den totale inntekten til innbyggerne i kommunen. Da denne oppgaven fokuserer på økonomisk ulikhet er det viktig å betrakte alle kilder til inntekt, inkludert kapitalinntekter. Dette er også en definisjon av inntekt som blir brukt

<sup>15</sup>SSB har kommunal inntektsdata frem til 2015, noe som gjør at jeg ikke kan inkludere datamateriale om pant og spill for 2016 og 2017 i denne oppgaven.



Figur 9: Brutto medianinntekt i alle kommuner fordelt fra lavest til høyest inntekt (tall fra 2015)

ellers i forskningslitteraturen om økonomiske ulikhet, blant annet av Piketty & Saez (2003, s.5). Price & Novak (2000, s.85) bruker gjennomsnittlig disponibel inntekt i husholdninger i tillegg til median inntekt, og finner at begge målene på inntekt gir like resultater. Et problem ved bruk av bruttoinntekter er at de ikke viser oss den disponible inntekten etter skatt. Med den progressive inntektsskatten vi har i Norge, vil vi få et utvalg med større ulikhet enn om vi hadde sett på netto medianinntekt.

For å se om variasjon i inntekten over tid påvirker etterspørselen etter lodd, er det nødvendig å inflasjonsjustere bruttoinntekten for å fange effekten av endringene i reell kjøpekraft. Jeg deflaterer derfor bruttoinntekten med årlig konsumprisindeks hvor jeg bruker utvalgsperiodens første år, 2012, som basisår. Prisen for et lodd i pantelotteriet holdes konstant på 50 øre gjennom hele utvalgsperioden, og har dermed en negativ realprisvekst. Gjennomsnittlig brutto medianinntekt i kommunene for hele utvalgsperioden er NOK 327 000, med et standard avvik på NOK 26 567. Høyeste medianinntekt finner vi i Bærum med NOK 420 700 (2015), mens laveste finner vi i Kautokeino med NOK 244800 (2012). Kommunenes gjennomsnittlige brutto medianinntekt har en realvekst fra NOK 320 700 i 2012 til 332 000 i 2015.

## 4.4 Kontrollvariable i økonometrisk modell

### Utdanning

Selv med inntekt som forklaringsvariabelen av størst interesse, er det viktig å inkludere andre variable som kan påvirke etterspørselen etter lotterispill i en økonometrisk studie av inntektselastisitet. Dette gjør vi blant annet for å unngå forventningsskjevne estimater som følge av *utelatt variabel skjevhet*. Dette vil bli diskutert nærmere i Kapittel 5.

Utdanningslengde blir av svært mange inkludert i regresjoner for inntektselastisiteten til pengespill. Dette blir ofte begrunnet med at utdanning har en signifikant effekt både på inntekt og tilbøyeligheten til å delta i pengespill. Psacharopoulos & Patrinos (2004, s.117) finner i sin omfattende litteraturgjennomgang av lønnsavkastningen til utdanning, at utdanningslengden har en positiv effekt på forventet lønn i samtlige land. Raaum et al. (1999, s.1) identifiserer samme effekt for norske arbeidstagere.

Utdanningsnivå blir også inkludert i studier av pengespill da lotterier ofte blir kritisert for å utnytte deler av befolkningen med lavere forståelse av sannsynlighet og forventningsverdier. Filosofen Voltaire skal etter sigende ha kalt lotteri for "en skattlegging av uvitenhet"<sup>16</sup> (Doroghazi 2009). Ved antagelsen om at en lengre utdanning gjør deg bedre rustet til å ta rasjonelle valg når det kommer til lotterier og pengespill, blir utdanningslengde ofte inkludert i empiriske studier av spillatferd.

Tidligere bidrag til forskningslitteraturen om determinanter av pengespilletterspørsel har en variert definisjon av befolkningens utdanningsnivå. Stranahan & Borg (1998, s.75) definerer utdanningsnivå som fullføring av highschool og påbegynt høyere utdanning, Price & Novak (2000, s.86) bruker andelen av befolkningen med fullført høyere utdanning som mål på kommunens utdanningsnivå, mens Hansen (1995, s.390) bruker tre utdanningsvariable gitt av andel av befolkningen i fylket som har mer enn 12 års skolegang, 12-15 års skolegang og over 16 års skolegang. Ved bruk av utdanningsdata fra norske kommuner måler Borge (2003, s.8) utdanningsnivået som andel av befolkningen over 16 år med grunnskole som høyeste fullførte utdanning.

De sprikende definisjonene av utdanningsnivå kan være noe av grunnen til at resultatene i disse studiene også er svært sprikende. Stranahan & Borg (1998) og Borge (2003) finner at folk med grunnskole som høyeste utdanning har høyere spilltilbøyelighet enn de med høyere

---

<sup>16</sup>Om Voltaire faktisk har sagt dette hersker det stor tvil, men utsagnet blir stadig trukket frem i forskningsartikler for å illustrere at deltagelse i lotterier er en tilsynelatende irrasjonell handling.



utdanningsnivåer, mens Hansen (1995) finner et positivt forhold mellom loddsalg og utdanningsnivå. Price & Novak (2000) finner et positivt forhold mellom utdanningsnivå og salget av lotto-kuponger, men et negativt forhold mellom utdanningsnivå og salget av skrapelodd. Dette indikerer at hvordan vi definerer kommunens utdanningsnivå vil være utslagsgivende for resultatet.

Dette bekrefter Hansen (1995) da hun finner at resultatene er svært sensitive til hvor mange utdanningsnivåer man inkluderer i den økonometriske modellen. Med inspirasjon Hansen (1995) inkluderer jeg tre utdanningsnivåer for å få et mer nyansert bilde på hvordan ulike utdanningsnivåer påvirker spilltilbøyeligheten. Jeg inkluderer en variabel for andelen av kommunens befolkning over 16 år uten høyere utdanning, en variabel for kort høyere utdanning<sup>17</sup> og en variabel for høyere utdanning lengre enn fire år. Dette gir en gjennomsnittlig andel kommunenes befolkning med kort høyere utdanning på 18,2% med et standardavvik på 3,7% og gjennomsnittlig andel med lengre høyere utdanning på 4,2% med et standardavvik på 2,5%.

## **Alderssammensetning**

Også alderssammensetningen i kommunene inkluderes i denne oppgaven for å unngå mulige utelatt variabel skjevheter. Bred økonomisk forskning viser at alder er en determinant for inntekt (Skirbekk 2004). Samtidig finner flere studier en signifikant sammenheng mellom alder og lotterispill. Både Clotfelter & Cook (1990, s.112) og Borge (2003, s.9) finner henholdsvis at tilbøyeligheten til lotterispill er høyest blant voksne i aldersgruppene på henholdsvis 25-65 og 15-45 år. Dette blir videre bekreftet av Beckert & Lutter (2008, s.483), som identifiserer at forholdet mellom lottodeltagelse og alder best blir representert ved en invertert U-form. Altså at unge og eldre spiller mindre enn middelaldrende.

For å finne ut hvordan alder påvirker etterspørselen etter lodd i pantelotteriet, samtidig som jeg kontrollerer de andre forklaringsvariablene for alderseffekter, inkluderer jeg variable for alderssammensetningen i kommunene. Disse variablene gir kommunens aldersfordeling ved å se hvor stor andel av befolkningen som innenfor aldersgruppene 0-17år, 18-49år, 50-66år og 67 år og oppover. Gjennomsnittlig alderssammensetning i kommunene var for utvalgsperioden 0-17 år: 21%, 18-49 år: 39%, 50-66 år: 22% og 67år+ 17%. Aldersfordelingen

---

<sup>17</sup>Med kort høyere utdannede regnes alle som har tatt mellom 120 og 240 studiepoeng eller har fullført en utdanning i universitets- og høyskolesystemet som er kortere enn fem år. (SSB, 2017)

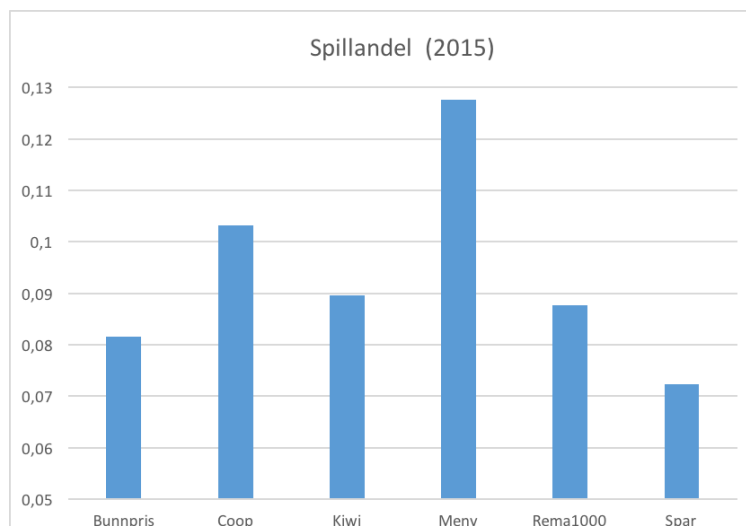
innad kommunene varierer naturlig nok svært lite da vi kun betrakter data for en tidsperiode på 4 år. Variasjonen mellom kommunene varierer i større grad. Her har vi at den gjennomsnittlige andelen av 18-49 åringer i kommunene på 39% har et standardavvik på 3,4%.

## Dagligvarekjede

Ved å inkludere informasjon om hvilken dagligvarekjede panteautomaten er plassert i, øker vi muligheten for å kunne kontrollere for forskjellige kundegrupper. Dagligvarekjedene varierer på pris, utvalg og beliggenhet. I en bestillingsrapport for forbrukerrådet klassifiserer Frode Alfnes og Arne Dulrud (2016, s.24) dagligvarekjedene følgende; *Nærbutikker* (Coop Marked, ICA Nær, Joker), *Lavpris* (CoopPrix, Coop Extra, Rimi, Kiwi, Bunnpris og Rema1000), *Bredsortiment* (Coop Mega, Coop Obs!, ICA Supermarked, ICA Maxi, Spar, Meny og Ultra). I tillegg varierer de forskjellige kjedene med egne leverandører og spesielle tilbudskampanjer. Med en slik variasjon i butikkene hvor panteautomatene er plassert, er det ønskelig å inkludere modellspesifikasjoner med dummyvariabler for dagligvarekjedene. På den måten kan vi se hvorvidt spilltilbøyeligheten varierer mellom kundegruppene til dagligvarekjedene. Etter som Lavik og Jacobsen (2015, s.18) finner at 83% av dagligvarehandler blir lagt til nærmeste butikk, er det nærliggende å anta at det er avstand og ikke butikktype som er avgjørende for valg av butikk. Dette sannsynliggjør at spilltilbøyeligheten ikke vil variere voldsomt mellom butikkjedene. Av Figur 10 ser vi imidlertid at det er en variasjon mellom dagligvarekjedene. Breddesortimentsbutikken Meny har klart høyest spillandel på 12,8% mens lavpriskjedene Bunnpris, Coop, Kiwi og Rema1000 har en spillandel mellom 8% og 11%. Hvorfor breddesortimentsbutikken Spar har den laveste spillandel på 7,2% er uvisst, og illustrerer at pantean delen ikke nødvendigvis er bestemt av butikktype.

## Sysselsetting

Sysselsettingens effekt på lotterier har blitt grundig studert i forskningslitteraturen. Ved arbeidsløshet og fattigdom kan pengespill representere eneste mulighet til å raskt kunne opparbeide seg en formue. Flere studier finner at arbeidsløsheten i et geografisk begrenset område har en positiv effekt på etterspørselen etter lotterispill. Mikesell (1994, s.169) finner i sin studie av amerikanske stater en positiv sammenheng mellom arbeidsløshet og lotterikonsum. Samtidig er det også blitt gjennomført studier hvor etterspørselen etter pengespill faller ved økt arbeidsledighet. Beckert & Lutter (2008, s.483) finner at sysselsatte deltar i statlige



Figur 10: Spillandel for ulike dagligvarekjeder (2015)

lotterier i større grad enn arbeidsledige. Dette forklares gjerne ved at rasjonelle risikooversere aktører skyr unna investeringer med negativ forventet avkastning, spesielt når man ikke lenger har lønnsinntekter.

Tall om sysselsettingen i denne oppgaven er hentet fra SSBs arbeidskraftundersøkelse og definert som andel sysselsatte i kommunen i aldersspennet 15-74 år. Gjennomsnittlig sysselsetting for utvalgsperioden er på 65,2% men varierer i stor grad fra kommune til kommune med et standardavvik på 14 prosentpoeng. Høyeste sysselsettingsgrad finner vi i Sandefjord i 2013 med 79,8%, mens laveste andel sysselsatte finner vi i Stor-Elvdal på 54,5%. Gjennomsnittlig observerer vi også en fallende grad av sysselsetting i utvalgsperioden fra 65,7% i 2012 til 63,2% i 2015.

## Bostedstype

Flere studier viser at befolkningstettheten rundt utsalgssteder for pengespill har en positiv effekt på konsumet av pengespill. Hansen (1995, s.394) finner at befolkningstettheten øker loddsalg per capita, mens Heavey (1978, s.424) finner at befolkningen som bor i indre by spiller mer enn befolkningen i forsteder. Befolkningstett og bostedstype blir gjerne inkludert som en forklaringsvariabel av to grunner. Først og fremst ønsker man å kontrollere for lotteriets tilgjengelighet. Bor man i byer eller tettsteder er utsalgssteder for pengespill nærmere potensielle kunder, noe som øker tilgjengeligheten og muligheten for konsum. Dette er spesielt relevant for pantelotteriet, da det kun er enkelte dagligvarekjeder som tilbyr dette lotteriet. Pantelotteriet blir dermed relativt utilgjengelig dersom man bor i områder langt unna

dagligvarebutikker som tilbyr pantelotteriet. Hansen (1995, s.393) argumenterer også for å inkludere bostedstype som en proxy-variabel for å få et mål på hvor urban en befolkning er. Dersom det eksisterer et skille mellom den urbane og den rurale befolkningens lotteripreferanser eller risikovillighet, kan inkludering av bostedstype i de empiriske estimeringene korrigere for dette.

Som mål på bostedstype inkluderer denne oppgaven et mål på andelen av befolkningen som bor i en by eller et tettsted i kommunene hvor panteautomatene er plassert. Et tettsted blir av SSB definert som *En hussamling hvor det bor minst 200 personer og avstanden mellom husene normalt ikke overstiger 50 meter*, og er å finne i alle kommunene i denne oppgavens datasett, utenom Flå. Gjennomsnittlig bor 82% av kommunenes befolkningen i tettbebygde strøk, med et standardavvik på 19%.

## Etnisitet

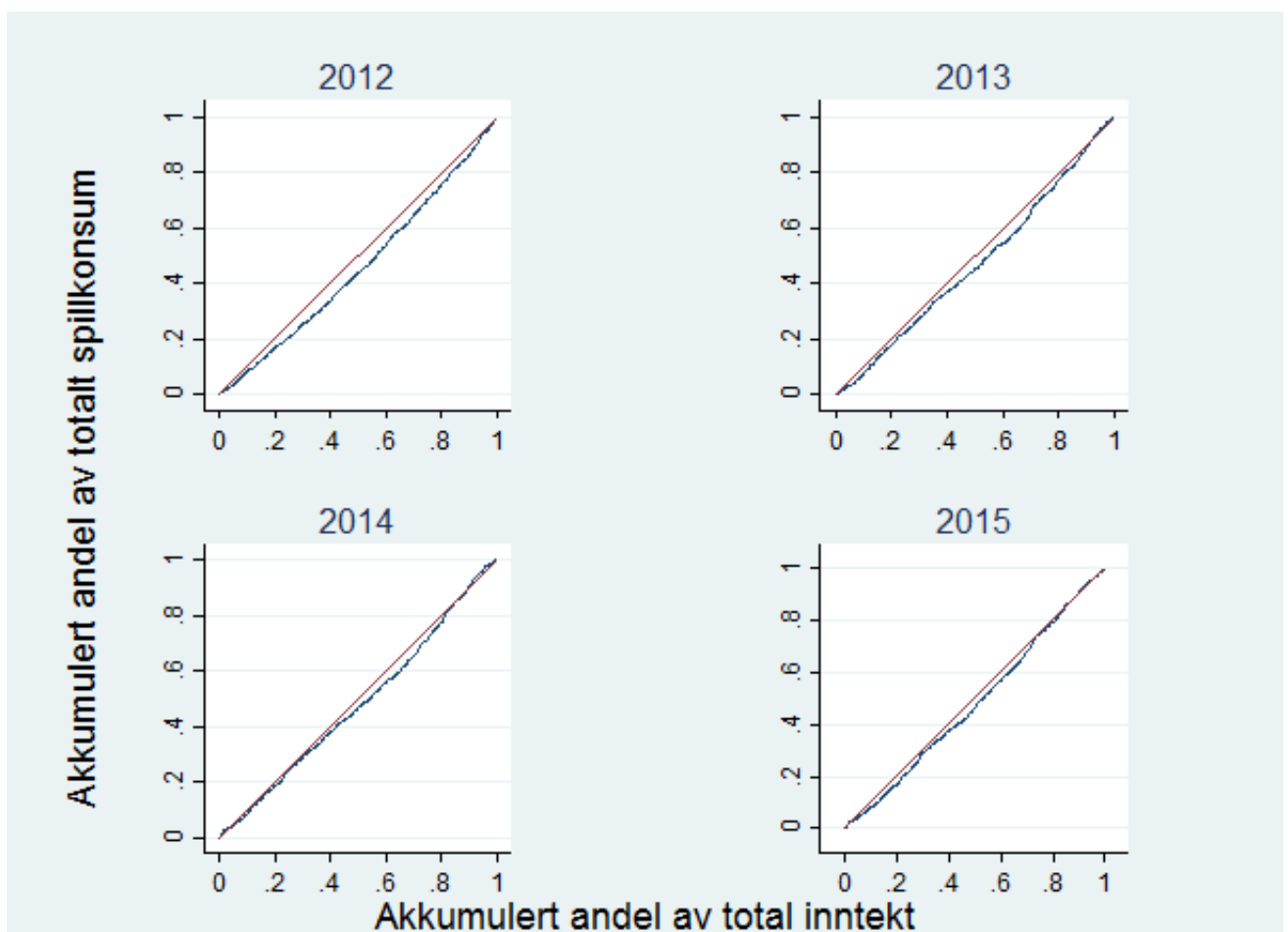
Store deler av litteraturen om lotterispill er utført ved bruk av datamateriale fra amerikanske spillmarkeder. Mye av denne forskningen og litteraturen lar seg overføre til det norske spillmarkedet, men enkelte forklaringsvariable er ikke like hensiktsmessige å anvende når vi betrakter det norske lotterimarkedet. Andel etniske minoriteter som er bosatt i kommunene er en slik variabel som ikke vil bli inkludert i denne oppgaven. Price & Novak (2000, s.88), Clotfelter & Cook (1987, s.19), Hansen (1995, s.389) og Stranahan & Borg (1998, s.79) viser alle at latinamerikanere og amerikanere av afrikansk herkomst har en større tilbøyelighet til å delta i pengespill enn majoritetsbefolkningen. Forskning gjort av Mikesell (1989, s.519) viser også at disse resultatene er spesielt fremtredende ved såkalte instant-spill hvor eventuell gevinst blir avdekket rett etter kjøp av lodd, som hos pantelotteriet. Grunnet store demografiske forskjeller lar minoritetseffekter fra USA seg vanskelig overføre til norske studier. Men med så tydelige minoritetseffekter fra andre land, er etnisitet en relevant forklaringsvariabel som ikke burde utelates. SSB fører ingen statistikk over personer fordelt etter religion, livssyn eller etnisitet,<sup>18</sup> så det er ikke praktisk gjennomførbart å inkludere dette som en forklaringsvariabel.

---

<sup>18</sup> SSB har en rikholdig statistikk om innvandrerbakgrunn og innvandringsland, men dette inkluderer ikke tilhørighet til etniske grupper.

## 4.5 Beregnede Suits-indeks

Suits-indeksen brukes som et deskriptiv mål på hvordan byrden av en skatt eller avgift fordeles over en befolkning. Som diskutert i Kapittel 3, viser de blå kurvene i Figur 11 forholdet mellom den akkumulerte andelen av total inntekt og den akkumulerte andelen av totalt spillkonsum for alle kommuner med pantelotteri, rangert etter brutto medianinntekt i kommunene. Disse kurvene deler mange av sine statistiske egenskaper med en Lorenz-kurve, og viser hvordan den implisitte skattebyrden av pantelotteriet treffer befolkningen i kommunene med pantelotteri.



Figur 11: Grafisk fremstilling av Suits-indeks for utvalgsårene 2012-2015.

Ved å betrakte den grafiske fremstillingen av Suits-indeksen i Figur 11, kan vi konstantere at mesteparten av Lorenz-kurven ligger under den diagonale linjen som indikerer en proporsjonal beskatning. Dette tilsier at pantelotteriet er progressivt beskattende, og at skattebyrden treffer kommuner med høy inntekt i større grad enn kommuner med lav. For å tallfeste størrelsen på skatteprogressiviteten, estimerer vi Suits-indeksen ved å anvende formlene gitt

i Likning (4), (7) og (8) i Kapittel 3.

**Tabell 1: Suits-indekser**

År	Suits-indeks	N
2012	0.083	264
2013	0.057	268
2014	0.038	300
2015	0.033	311
Gj.snitt	0.053	

Som Figur 11, indikerer også de tallfestede Suits-indeksene i Tabell 1 en progressiv beskatning for alle utvalgsår. Vi ser her at Suits-indeksene er positive gjennom hele utvalgsperioden, med indeksverdier i spennet 0.033-0.083. En gjennomsnittlig indeksverdi på 0.053 skiller seg fra de fleste andre studier av lotteri og pengespill ved å være positiv. En litteraturgjennomgang utført av Perez & Humphreys (2013, s.930) oppsummerer Suits-indekser for 11 studier av lotterier og pengespill, og finner at samtlige har en negativ Suits-indeks. Resultatene varierer mellom de ulike studiene, og ligger i spennet mellom -0.46 og -0.03. Studier gjort med data på fylke/kommune-nivå har systematisk en Suits-indeks nærmere 0.

Av studier utført med områdedata, får Price & Novak (2000, s.87), Hansen (1995) og Borge (2003) Suits-indekser på henholdsvis -0.12, -0.095 og -0.096. Price & Novak (2000) og Hansen (1995) argumenterer for at bruken av områdedata vanner ut variasjonen mellom individer og at dette virker dempende på den kalkulerede Suits-indeksen. Selv om pantelotteriet skiller seg fra sammenliknbare studier ved å ha positive Suits-indekser, føyer disse resultatene seg inn i rekken av områdedata-studier med Suits-indekser relativt nære 0.

Suits-indeksene i Tabell 1 viser en bevegelse mot en mer proporsjonal beskatning gjennom utvalgsperioden. Suits-indeksen faller fra 0.083 i 2012 til 0.033 i 2015. Når man betrakter endringer i fordelingseffekter av skatter over tid, er det viktig å være bevisst på at endringer i Suits-indeksen ikke nødvendigvis er forårsaket av en vridning i skattebelastningen av ulike inntektsgrupper. Suits-indeksen vil også bli endret dersom den underliggende inntektsfordelingen endres. Tall fra Statistisk sentralbyrå (2018) viser at inntektsulikheten, målt ved en Gini-koeffisient, økte fra 0.249 i 2012 til 0.271 i 2015. Denne økte ulikheten kan ha bidratt til at pantelotteriet som en beskatning treffer mer proporsjonalt utover i utvalgsperioden.

For å se om resultatet fra Suits-indeksene er sensitivt til hvordan vi definerer inntekten, kalkulerer vi Suits-indeksene med brutto gjennomsnittsinntekt i kommunene som inntektsmål. Tabell 2 viser at den gjennomsnittlige Suits-indeksen for hele utvalgsperioden øker marginalt fra 0.053 til 0.056, mens utviklingen over tid er noe mer markant når vi bruker gjen-

nomsnittsinntekt. Resultatene i Tabell 2 antyder at valget av mål på inntekt ikke er utslagsgivende for de kalkulerte Suits-indeksene.

**Tabell 2: Suits-indeks med ulike inntektsmål**

<b>År</b>	<b>Gjennomsnittsinntekt</b>	<b>Medianinntekt</b>	<b>N</b>
2012	0.094	0.083	264
2013	0.061	0.057	268
2014	0.051	0.038	300
2015	0.018	0.033	311
Gj.snitt	0.056	0.053	

## 5 Økonometrisk rammeverk for estimering av inntektselastisitet

Det tilgjengelige datamaterialet gjør at vi kan konstruere et ubalansert paneldatasett, hvor informasjon fra de enkelte panteautomatene kobles sammen med regional informasjon fra kommunene de er plassert i. Hver panteautomat blir identifisert som en tverrsnittenheter som vi følger over utvalgsperioden på 4 år. Ettersom pantelotteriet gradvis blir introdusert i de ulike dagligvarekjedene, blir ikke alle panteautomater observert gjennom hele utvalgsperioden. Dette gir oss et ubalansert paneldatasett (Wooldridge 2013, s.373). Paneldata gir oss muligheter til å ta i bruk en rekke modellspesifikasjoner i vår empiriske estimering av inntektselastisiteten til pantelotteriet.

For å identifisere hvilke modellspesifikasjoner det er mest hensiktsmessig å ta i bruk i den økonometriske analysen, er det nødvendig å diskutere hvilke økonometriske utfordringer dette paneldatasettet byr på. Diskusjonen vil legge grunnlaget for hvilke modellspesifikasjoner som gir mest presise og troverdige resultater og som dermed bør inkluderes i den økonometriske analysen.

For å belyse potensielle statistiske problemer og fallgruver datasettet kan by på, tar vi utgangspunkt i et enkelt standardoppsett for paneldata.

$$y_{it} = \beta X_{it} + u_{it} \quad (10)$$

Her er  $y_{it}$  den avhengige variabelen og vil i den økonometriske analysen gi oss andelen av pantepengene som går til lotterispill.  $X_{it}$  er en matrise av våre forklaringsvariable og  $\beta$  angir den tilhørende koeffisientvektoren. Paneldatamodellen gir oss variasjon over både enheter og tid. Fotskrift  $i$  gir oss hvilken tverrsnittenheter observasjonen tilhører, mens  $t$  angir tidsenheten. Panteautomatene fungerer som tverrsnittenheter i denne oppgavens paneldatasett, mens tidsenhetene blir oppgitt som år.

$u_{it}$  er et stokastisk restledd som varierer både over tid og mellom tverrsnittenheter. Vi kan videre utvide det stokastiske restleddet til to komponenter.

$$u_{it} = \alpha_i + \epsilon_{it} \quad (11)$$

Det dekomponerte restleddet i Likning (11) består nå av et enhetsspesifikt restledd,  $\alpha_i$ ,



og et idiosynkratisk restledd,  $\epsilon_{it}$ . Det enhetsspesifikke restleddet fanger opp all tidsinvariant uobservert variasjon mellom panteautomatene som påvirker spilltilbøyeligheten. Slike tidsinvariante variable kan være størrelse på butikk, plassering av panteautomat i butikk eller butikkens beliggenhet. Det idiosynkratiske restleddet fanger opp all uobservert variasjon mellom panteautomatene og variasjon over tid.

Vi antar videre at restleddskomponentene oppfyller følgende standardantagelser om stokastisitet og homoskedastisitet:

$$E(\epsilon_{it}|x_{it}) = 0 \quad (\text{I})$$

$$E(\alpha_i|x_i) = 0 \quad (\text{II})$$

$$E(\epsilon_{it}, \epsilon_{js}) = \begin{cases} \sigma_\epsilon^2 & \text{for } i=j, t=s \\ 0 & \text{ellers} \end{cases} \quad (\text{III})$$

$$E(\alpha_i, \alpha_j) = \begin{cases} \sigma_\alpha^2 & \text{for } i=j \\ 0 & \text{ellers} \end{cases} \quad (\text{IV})$$

$$E(\epsilon_{it}|\alpha_i|x_{it}) = 0 \quad (\text{V})$$

Antagelse I og II gir oss at restleddskomponentene må være ukorrelerte med de inkluderte forklaringsvariable, mens antagelse III og IV forutsetter at restleddskomponentene er homoskedastiske, med konstant varians over alle observasjoner og fravær av seriekorrelasjon. V sier oss at det enhetsspesifikke restleddet må være ukorrelert med det idiosynkratiske restleddet.

Vi tar utgangspunkt i en regresjon ved bruk av *Ordinary Least Squares*<sup>19</sup> (OLS) som estimeringsmetode for å belyse eventuelle økonomiske problemer vi må forholde oss til ved bruk av paneldatasettet vi har konstruert. OLS vil gi oss konsistente og forventningsrette estimatorene dersom Gauss-Markov-forutsetningene for OLS er oppfylt Wooldridge (2013, s.79).

1. Modellen er lineær i parameterne
2. Utvalget er tilfeldig trukket
3. Ingen perfekt kollinearitet i forklaringsvariablene

<sup>19</sup>Også kjent som Minste Kvadraters Metode (MKM)

4. Restleddet er ukorrelert med de inkluderte forklaringsvariable,  $E(u_{it}|X_{it}) = 0$ .

Dersom restleddsantagelsene I-V og forutsetningene 1-4 er oppfylt, vil bruk av enkel OLS gi oss konsistente og forventningsrette koeffisienter. Bruk av OLS-estimatorer vil da utnytte både variasjon mellom tverrsnittene og variasjon innad i tverrsnittene over tid.

## 5.1 Økonometriske utfordringer

Dersom forutsetning I eller II blir brutt, vil ikke OLS lenger gi oss konsistente parameterestimatorer. Dette kapitlet vil se på hvordan oppgavens paneldatasett oppfyller de gitte forutsetningene, og hvordan vi kan takle eventuelle brudd på disse.

## 5.2 Endogene forklaringsvariable

Et brudd på antagelsen om strengt eksogene forklaringsvariable,  $E(u_{it}|X_{it}) \neq 0$ , vil gi oss inkonsistente forventningsskjevne estimatorer. Wooldridge (2013) trekker frem tre mulige problemer med datasett som kan føre til at denne eksogenitetsbetingelsen blir brutt, *målefeil*, *utelatt variabel problem* og *simultanitet*.

### Målefeil

Et datasett er utsatt for målefeil når dets observasjoner avviker fra den faktiske verdien til de inkluderte variablene. Målefeil er i mange datasett ikke til å unngå, men trenger ikke nødvendigvis å representere et stort endogenitetsproblem.

Er den avhengige variabelen heftet med en målefeil som ikke er systematisk koblet til forklaringsvariablene i modellen, vil Gauss-Markov-forutsetningene fortsatt være oppfylt, og OLS-estimatorene vil være forventningsrette og konsistente. Dersom målefeilene er av en større art, vil dette resultere i en høyere varians og mindre presise parameterestimer ved bruk av OLS (Wooldridge 2013, s.309).

Spillandelen, som er denne analysens avhengige variabel kan være heftet med enkelte målefeil da paneldatasettet består av årlige tverrsnitt. Enkelte panteautomater kan ha blitt installert sent på året, slik at de får svært få pantesesjoner i løpet av det året vi betrakter. Små utvalg kan gi et feilaktig bilde av spillandelen da man ikke har nok observasjoner til å regne ut et troverdig spillandel. Enkelte panteautomater med svært få pantesesjoner hadde en spil-

landel lik 1. Av den grunn har jeg fjernet alle butikker med færre enn 100 spillsesjoner<sup>20</sup>. Vi kan heller ikke identifisere i hvilken måned panteautomatene ble installert i butikkene. Dette gjør at vi med sesongeffekter som økt spilling i desember risikerer å få kunstige høye spillandeler fra panteautomater som kun viser data for desember måned det året den blir installert. Det er ingen grunn til å tro at måneden automaten blir installert i butikken er korrelert med inkluderte forklaringsvariable i analysen, og at dette dermed ikke vil gi oss en systematisk målefeil og brudd på forutsetningen om eksogene forklaringsvariable.

Målefeil av forklaringsvariable representerer et større problem da målefeil i forklaringsvariabelen svekker korrelasjonen mellom forklaringsvariabelen og den avhengige variabelen. Gitt at målefeilen er ukorrelert med den aktuelle forklaringsvariabelen, vil koeffisienten bli forventningsskjev mot null (Wooldridge 2013, s.311). Denne skjevheten vil øke med størrelsen på målefeilen.

### Utelatt variabel-skjevhet

Dersom en utelatt variabel er korrelert med både den avhengige variabelen og en eller flere av de inkluderte forklaringsvariablene, vil vi stå ovenfor et utelatt variabel-problem. De inkluderte variablene som er korrelert med den relevante utelatte variabelen vil med dette være korrelert med restleddet, noe som gjør at forutsetning 4.,  $E(u_{it}|X_{it}) = 0$ , vil være brutt. Brudd på denne forutsetningen vil gi oss forventningsskjeve estimatorer, da de inkluderte variablene også vil fange opp noe av effekten til den utelatte variabelen. Vi vil ved bruk av OLS altså kun ha forventningsrette estimatorer dersom det ikke er utelatt noen relevante variabler som er korrelert men noen av de inkluderte forklaringsvariable.

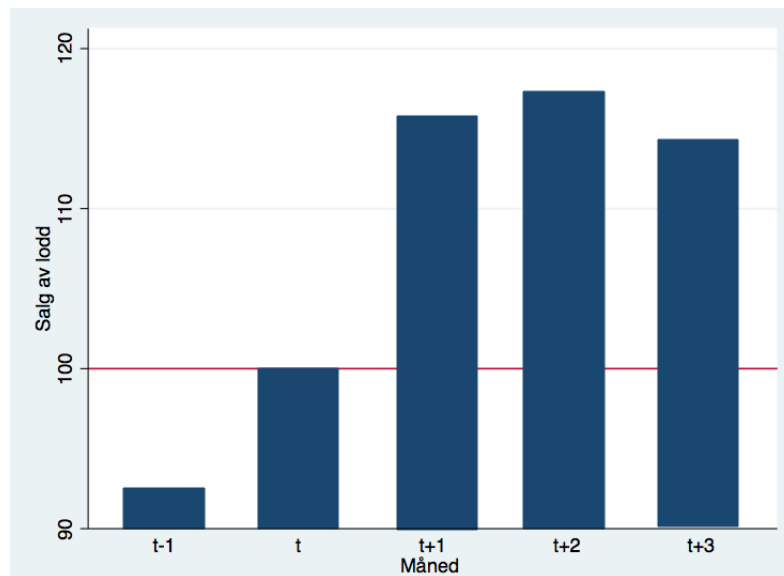
Det er derfor et ideal å inkludere alle relevante forklaringsvariabler som også korrelerer med de andre variable i modellspesifikasjonen vår. I praksis er dette en umulig oppgave og vi må derfor anta at OLS på paneldatasettet vårt vil være heftet med muligheten for en utelatt variabel-skjevhet.

For å se på noen eksempler som illustrerer dette problemet, tar vi utgangspunkt i et dekomponert restledd med en enhetsspesifikk og en idiosynkratisk komponent,  $u_{it} = \alpha_i + \epsilon_{it}$ .

Den idiosynkratiske restleddskomponenten,  $\epsilon_{it}$ , inneholder uobserverte variable som variere både mellom tid og enheter. En slik variabel kan være måten lotterikunder oppda-

---

<sup>20</sup>60 butikker hadde under 100 spillsesjoner i løpet av et år, gjennomsnittlig antall årlige spillsesjoner for utvalget er 3072,8.



Figur 12: Endringer i loddsalg i butikker som har solgt et vinnerlodd

terer sin subjektive oppfatning av sannsynligheten for å kjøpe et vinnerlodd i forskjellige butikker. Stanfield & Kearney (2008) har identifisert sette som *Lucky Stores*-effekten. Her finner de at butikker som solgt et vinnerlodd for en topp-premie opplever en 12-38% økning i loddsalg de følgende ukene, og at denne effekten er avtagende men gjeldene i hele 40 uker. I 2016 ble 9 av 11 vinnere av pantelotteriets milliongevinst omtalt i lokale og landsdekkende nyhetsoppslag. Ved å se på gjennomsnittlig spillandel i butikker som har solgt et lodd med millionpremie, kan vi se at denne effekten også gjør seg gjeldende for pantelotteriet. Figur 12 viser hvordan gjennomsnittlig salg i vinnerbutikkene i 2016 utvikler seg fra måneden vinnerloddet ble solgt,  $t$ , og tre måneder frem. Her øker gjennomsnittlig spillandel med 17,3% to måneder etter salget av et vinnerlodd, for så å avta noe etter tre måneder. Vi ser også at det gjennomsnittlige salget i måned  $t$  er adskillig høyere enn måned  $t-1$ . Dette skyldes at vinnerloddet i enkelte butikker kan ha blitt kjøpt tidlig i måned  $t$ , og at salget i denne måneden da også vil bli påvirket av lucky stores-effekten. Når deltagere i pantelotteriet tilsynelatende trekker mot butikker som har solgt et vinnerlodd, vil dette gi oss et seleksjonsproblem dersom denne effekten ikke treffer jevnt over befolkningen.

Guryan & Kearney (2005, s.467) viser at lucky stores-effekten er mest fremtredende i befolkninger med høyere andel av demografisk svakerestilte grupper som andel av befolkningen som lever i fattigdom og andel uten fullført videregående skole. Det er rimelig å anta at lucky stores-effekten vil være korrelert med utdanningsnivået til kommunen, som gjør at vi har en utlatt variabel som både er korrelert med den avhengige variable og en av model-

lens forklaringsvariable. Størrelsen på dette problemet må derimot diskuteres. En midlertidig salgsøkning hos rundt 10 vinnerautomater per år representerer en svært liten variasjon i et datasett med flere tusen automater, noe som tilsier at en utelatt variabel-skjevhet fra vinnerbutikker ikke representerer et problem for denne analysen. Samtidig illustrerer dette eksempelet hvordan en uobservert variabel kan være korrelert med både de inkluderte forklaringsvariable og den avhengige variabelen.

Den enhetsfaste restleddskomponenten,  $\alpha_i$ , gir oss utelatte og uobserverte faktorer som påvirker den avhengige variabelen og som varierer mellom tversnittene, men er faste over tid. Dette kan være egenskaper som størrelsen til butikken, plassering av panteautomat eller kommunefaste effekter. Det er meget sannsynlig at denne oppgavens datasett er heftet med uobservert heterogenitet mellom panteautomatene og kommunene, og det er grunn til å mistenke at antagelse II,  $E(\alpha_i|x_i) = 0$ , ikke vil holde. Dette gjør at OLS vil gi oss inkonsistente estimatorer. Ideelt sett bør vi da ta i bruk alternative estimeringsmetoder som eliminerer det enhetsspesifikke restleddet, samtidig som vi beholder nok variasjon til å estimere konsistente estimatorer. Jeg kommer tilbake til diskusjon av dette problemet senere.

## Simultanitet

Endogenitet i forklaringsvariablene kan også oppstå dersom en eller flere av forklaringsvariable bestemmes simultant som den avhengige variabelen. Når forklaringsvariabelen bestemmes simultant med den avhengige variabelen, vil den bli korrelert med modellens restledd og vi får igjen at antagelse II vil bli brutt,  $E(\alpha_i|x_i) \neq 0$ . Dette problemet kan oppstå ved at vi har en toveis kausal effekt, altså at både forklaringsvariabelen og den avhengige variabelen påvirker hverandre.

I et vanlig lotteri, hvor premiestørrelsen avhenger av loddsalget kan vi finne et slikt simultanitetsproblem. Her gir økt salg av lodd en større jackpot, samtidig har studier vist at størrelsen på jackpotten øker loddsalget (Forrest et al. 2002, s.495). Dette har blitt lansert som et mulig simultanitetsproblem for å identifisere drivere av lotterietterspørselen (Garrett & Sobel 2004). Analysen av pantelotteriet unngår dette problemet ved å ha en fast premiestørrelse gjennom hele utvalgsperioden.

En mulig kilde til simultanitet kan oppstå ved at en andel av overskuddet til pantelotteriet går til Røde Kors sitt lokallag der panteautomaten er plassert. Disse lokallagene har flere samfunnsnyttige formål, som integrering av flyktninger eller sosial inkludering av tidligere

straffedømte. Om dette arbeidet er utbredt og skjer i stor skala, er det teoretisk mulig at en økt spillandel på panteautomater kan bidra til økt sysselsetting. Da studier viser at sysselsetting har en signifikant effekt på etterspørselen etter lotterier (Beckert & Lutter 2008, s.483), kan det oppstå en toveis kausal effekt, som vil gi oss et simultanitetsproblem. At pantelotteriets overskudd skal få reelle konsekvenser for sysselsettingen i kommunene er lite troverdig, men er et eksempel på at det er teoretiske kilder til simultanitet i denne oppgavens datasett.

Generelt er de inkluderte forklaringsvariable i denne oppgaven lite volatile og gjenspeiler dyptgående prosesser og strømninger i samfunnet. Det er lite trolig at spillandelen vil ha en *gjensidig kausalitet* med demografien, utdanningslengden eller inntekten i de ulike kommunene. Simultanitet representerer derfor ingen stor risiko for brudd på forutsetningen om at restleddet må være ukorrelert med de inkluderte forklaringsvariable.

## Multikollinearitet

For at Gauss-Markov forutsetningen om ingen perfekt kollinearitet mellom forklaringsvariablene skal holde, kan ingen forklaringsvariable bli gitt nøyaktig av en annen forklaringsvariabel eller en lineær kombinasjon av de andre forklaringsvariablene. Er dette tilfellet, vil det bli umulig å skille hvilke av de perfekt korrelerte forklaringsvariablene som påvirker den avhengige variabelen ettersom de alltid vil opptre sammen. Også korrelasjon mellom forklaringsvariablene som ikke er perfekt kan svekke modelens produksjonsevne ved å vanskeliggjøre identifikasjon av de partielle effektene til de korrelerte forklaringsvariablene. Dette gjør at jo mer korrelert forklaringsvariablene er, dess høyere vil variansen til de respektive estimerte effektene av forklaringsvariablene bli (Wooldridge 2013, s.90). Er korrelasjonen mellom to eller flere forklaringsvariable tilstrekkelig høy <sup>21</sup>, vil vi få et problem med multikollinearitet. Vi får altså ubrukelige estimater om vi har perfekt kollinearitet mellom forklaringsvariablene og mindre presise estimater dersom det er høy korrelasjon mellom dem. Multikollinearitet i datasettet bryter ingen av forutsetningene for forventningsrette og konsistente estimater, men kan likevel by på problemer dersom høy korrelasjon mellom forklaringsvariablene gjør det umulig å få presise estimater. Vi kan se av korrelasjonsmatrisen i Tabell 11 i appendikset at enkelte av forklaringsvariablene er innbyrdes høyt korrelerte. Dette gjelder spesielt de variablene som er gjensidig ekskluderende, som aldersgrupper og utdan-

<sup>21</sup>Wooldridge (2013, s.91) bedyrer at det ikke er noen klar definert grense for hvor høy forklaringskraften,  $R^2$ , må være for at multikollineariteten blir problematisk og fremhever at størrelsen på datasettet kan kompensere for høy grad av multikollinearitet.

ningsgrupper. Viktigst er det at interessevariabelen, *Inntekt*, ikke er høyere korrelert enn 0.75 med de andre forklaringsvariabelene. Dette tyder på at multikollinearitet ikke representerer et stort problem for den estimerte inntektselastisiteten i denne oppgaven. Om høyt korrelerte variable representerer et problem ved å påvise signifikans, kan vi utelate en av de strekt korrelerte variablene. Dette gir derimot et nytt problem da vi med å utelate en muligens relevant variabel som er korrelert med en eller flere forklaringsvariable skaper et utelatt variabel problem. Gitt at multikollineariteten ikke gir oss en problematisk høy varians vil det være naturlig å inkludere alle forklaringsvariable i en generell modell. Vi kan estimere forenklinger hvor vi utelukker de høyest korrelerte variablene for så å teste om disse forenklingene er gyldige.

### Seleksjonsskjevheter fra kundeseleksjon

Seleksjon av kunder til de ulike butikkene kan gi oss forventningsskjevne estimatører.

Ved å bruke makrodata fra kommunene hvor panteautomatene er plassert, antar vi også at denne informasjonen representerer kundemassen i panteautomatenes respektive butikker. Dette datasettet fanger ikke opp informasjon om kundegruppen til de spesifikke panteautomatene, og vi står dermed ovenfor et seleksjonsproblem. De ulike dagligvarekjedene hvor panteautomatene er plassert retter seg mot ulike markedssegmenter og det er rimelig å anta at de tiltrekker seg ulike kundegrupper. Om mer velstående kunder unngår lavpriskjeder til fordel for butikker med breddesortiment blir ikke fanget opp av dette datasettet. Figur 10 viser at den avhengige variabelen, spillandelen, varierer mellom dagligvarekjedene, hvor Meny har klart høyest spillandel. Vi vil derfor inkludere hvilken dagligvarekjede butikkene hører til i modellspesifikasjonen for å kunne kontrollere for ulike kundegrupper, slik at vi kan håndtere det potensielle seleksjonsproblemet.

### Ubalansert paneldatasett

Pantelotteriet blir gradvis implementert i flere butikker gjennom utvalgsperioden, med en økning fra 1897 panteautomater i 2012 til 2357 i 2015. Totalt inneholder datasettet observasjoner fra 2851 panteautomater, som gjennomsnittlig observeres over 2,9 år. Dette gjør at vi følger tverrsnittene over tidsperioder av ulik lengde, og vi får et *ubalansert datasett*. Her er 1371 panteautomater blir fulgt over 4 år, 291 over 3 år, 784 over 2 år og 405 over ett år. Wooldridge (2013, s.473) advarer mot bruk av ubalanserte paneldatasett dersom grunnen til

manglende data er korrelert med det idiosynkratiske restleddet. I tilfellet med pantelotteriet er avgjørelsen om hvilke butikker som innfører pantelotteriet i stor grad eksogent bestemt ved at hele butikkjeder innfører pantelotteriet samtidig. Dette gjør at avgjørelsen ikke ligger hos de enkelte butikkene, noe som reduserer sjansen for et selektert utvalg. Samtidig vil vi også kontrollere for dagligvarekjede i den økonometriske modellen.

### **Lite variasjon i forklaringsvariable**

Som vi så i presentasjonen av denne oppgavens forklaringsvariable, er det en relativt stor variasjon mellom kommunenes demografiske karakteristika som inntekt, alderssammensetning og utdanningsnivå. Dette gjør at empiriske modellspesifikasjoner som utnytter variasjon mellom tverrsnittenhetene vil ha gode forutsetninger for presise estimatorer, så lenge restleddsantagelsene I-V er oppfylt. Som diskusjonen av utelatt variabel-skjevhet, gjør uobserverte tidsinvariante variable gjøre at restleddsantagelsen,  $E(\alpha_i|x_i) = 0$ , trolig er brutt, og at vi bør vurdere økonometriske modellspesifikasjoner som transformerer bort uobserverbar variasjon mellom tverrsnittenhetene. Eksempler på modeller som anvender enhetsfaste effekter er *Fixed Effects*- (FE) eller *First Difference*-modeller (FD). Om vi transformerer bort all variasjon mellom tverrsnittenhetene vil reduksjonen av informasjon trekke i retning av mindre presise estimater. Samtidig blir vi helt avhengige av tilstrekkelig variasjon innad i tverrsnittenhetene (Wooldridge 2013, s.457). Dette kan fort bli et problem da de fleste inkluderte forklaringsvariable varierer svært lite av natur. Kommunenes alderssammensetning, utdanningsnivå og medianinntekt varierer lite fra år til år. Det er heller ingen store økonomiske sjokk gjennom utvalgsperioden, som kunne ha gitt interessevariabelen medianinntekt en større variasjon innad i kommunene. Dette reflekteres i en beskjeden gjennomsnittlig realvekst i kommunenes brutto medianinntekt fra NOK 320 700 i 2012, til NOK 332 000 i 2015. Liten variasjon innad i tverrsnittenhetene gjør at vi mister en svært stor andel av paneldatasettets totale informasjon om vi transformerer bort variasjonen mellom tverrsnittenhetene. Denne sårbarheten blir spesielt fremtredende da vi kun har fire tidsperioder.

### **Seriekorrelerte restledd**

Selv om restleddsantagelsen om et ukorrelert enhetsspesifikt restledd holder, vil tradisjonell OLS-estimering likevel være problematisk. Dette kommer av at den enhetsspesifikke restleddskomponenten,  $\alpha_i$ , av natur vil være konstant gjennom alle perioder i et paneldatasett.



Dette gir oss et seriekorrelert restledd. Seriekorrelasjon i restleddet gjør at vi ikke er i stand til å finne konsistente standardavvik ved bruk av OLS. Uten konsistente standardavvik vil vi ikke kunne teste koeffisientene for signifikans.

Dette kan vi se gjennom varians-kovarians-matrisa til hver tverrsnittsenhet, som er gitt;

$$\Omega_i = E(u_{it}u'_{it}) = \begin{bmatrix} \sigma_\alpha^2 + \sigma_\epsilon^2 & \sigma_\epsilon^2 & \dots & \sigma_\epsilon^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_\epsilon^2 & \dots & \sigma_\alpha^2 \\ \dots & \dots & \dots & \dots \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \dots & \sigma_\alpha^2 + \sigma_\epsilon^2 \end{bmatrix}_{T \times T}$$

Hvor  $\sigma_\alpha^2$  er den enhetsspesifikke restleddvariansen og  $\sigma_\epsilon^2$  er den stokastiske restleddvariansen.

Vi kan se at samtlige elementer utenfor hoveddiagonalen til  $\Omega_i$  er ulike 0, og at denne varians-kovarians-matrisa dermed ikke har en standard struktur. Ringvirkningene av dette kan vi se om vi betrakter den komplette varians-kovarians-matrisa til datasettet.

$$\Omega = \Omega_i \otimes I = \begin{bmatrix} \Omega_i & 0 & \dots & 0 \\ 0 & \Omega_i & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \Omega_i \end{bmatrix}$$

Seriekorrelasjon i  $\Omega_i$  gjør at den komplette varians-kovarians-matrisa vil anta en form som er ulik  $I * konstant$ , som tilsier at vi vil få inkonsistente standardavvik. Dette gjør at vi bør bruke *cluster-robuste* standardavvik for å estimere konsistente standardavvik som kan brukes til å påvise inferens.

## Random Effects/GLS-estimator

Bruker vi først cluster-robuste standardavvik, vil vi ikke få effisiente OLS-estimatorer, etter som det vil eksistere GLS-estimatorer (*Generalized Least Squares*) med høyere presisjonsnivå. GLS-estimatoren er gitt

$$\hat{\beta}^{GLS} = (X' \Omega^{-1} X)^{-1} (X' \Omega^{-1} Y) \quad (12)$$

hvor  $\hat{\beta}^{GLS}$  er den komplette koeffisientvektoren, X er komplett matrise av forklaringsvariable og Y er en vektor for den avhengige variabelen. Estimatoren i Likning (12) implemen-

teres ved å transformere den generelle modellen i Likning (10) til en form med bedre restleddegenskaper. Denne estimeringsmetoden går også under navnet *Random Effects* (RE).

$$y_{it} - (1 - \theta)\bar{y}_i = \beta(X_{it} - (1 - \theta)\bar{X}_i) + (u_{it} - \theta\bar{u}_i) \quad (13)$$

Her er  $\bar{y}_i$ ,  $\bar{X}_i$  og  $\bar{u}_i$  enhetsgjennomsnitt over tid, og  $\theta$  er en transformasjonsparameter som angir en fast andel av enhetsgjennomsnittet som vil bli trukket fra observasjonene.

Dette gjøres ved at  $\theta$  blir gitt av størrelsesforholdet mellom den enhetsspesifikke restleddvariansen,  $\sigma_\alpha^2$ , og den stokastiske restleddvariansen,  $\sigma_\epsilon^2$ .

$$\theta = \sqrt{\frac{\sigma_\alpha^2}{\sigma_\epsilon^2 + T * \sigma_\alpha^2}} \quad (14)$$

Dette betyr at størrelsen til transformasjonsparameteren,  $\theta$ , blir gitt av hvor stor trussel  $\sigma_\alpha^2$  utgjør for et seriekorrelert restledd. Er  $\sigma_\alpha^2$  høy vil vi ha et problem med seriekorrelasjon ettersom den enhetsspesifikke restleddkomponenten,  $\alpha_i$ , vil være konstant over alle perioder. Av Likning (14) ser vi at dette problemet øker med antall perioder,  $T$ .  $\theta$  vil dermed bli mindre når  $T * \sigma_\alpha^2$  øker, og vi får en estimator som i større grad vektlegger variasjon innad i tverrsnittene frem for variasjon mellom tverrsnittene. Dersom det ikke finnes enhetsspesifikke restledd,  $T * \sigma_\alpha^2 = 0$ , vil transformasjonsparameteren  $\theta = 1$  gjøre at RE-estimatoren blir lik OLS-estimatoren. Motsatt, om  $T * \sigma_\alpha^2 \approx \infty$ , vil  $\theta = 0$  gi oss en estimator renser bort all variasjon mellom tverrsnittene. Dette vil gi oss like resultater som ved bruk av *Fixed Effects* som estimeringsmetode.

GLS-transformasjonen i Likning (13) vil eliminere seriekorrelasjonen i restleddet. OLS på Likning (13) vil, under antagelsene I-V, gi oss konsistente og effisiente estimatører (Wooldridge 2013, s.475). RE-estimatoren får dermed konsistente og troverdige standardavvik samtidig som den utnytter variasjon fra både tids- og tverrsnitts-dimensjonen. Dette gjør *Random Effects* er en svært gunstig estimeringsmetode, så lenge antagelsen om et strengt eksogene restledd holder,  $E(\alpha_i | x_i) = 0$

Wooldridge (2013, s.475) trekker frem at GLS-prosedyren fungerer best for datasett med et høyt antall tverrsnittenheter og relativt få perioder. Dette er karakteristikk som også passer denne oppgavens datasett.

### Fixed Effects-estimator

Dersom antagelsen om strengt eksogene restledd blir brutt,  $E(\alpha_i|x_i) \neq 0$ , er *Fixed Effects* en estimeringsmetode som fortsatt vil gi oss forventningsrette estimatorer. Vi kan fjerne det korrelert enhetsspesifikke restleddet,  $\alpha_i$ , ved å innføre en dummyvariabel for hver tverrsnit-tenhet. Denne dummyvariabelen vil fange opp alle enhetsfaste effekter, inkludert  $\alpha_i$ . Ved å withintransformere Likning 10, subtraherer vi alle elementene med sine respektive enhetsgjennomsnitt. Withintransformasjonen fjerner da de enhetsfaste effektene og transformerer bort  $\alpha_i$ . Dette etterlater oss med et restledd som er ukorrelert med de inkluderte forklaringsvariable. En slik withintransformasjon vises i Likning 15 (Wooldridge 2013, s.467).

$$y_{it} - \bar{y}_i = \beta(X_{it} - \bar{X}_i) + (u_{it} - \bar{u}_i) \quad (15)$$

FE-estimatorene vil være konsistente så lenge den idiosynkratiske restleddskomponenten er ukorrelert med de inkluderte forklaringsvariable,  $E(\epsilon_i|x_i) = 0$ . Withintransformasjonen vil, samtidig som den transformerer bort det enhetsspesifikke restleddet, også rense bort all variasjon mellom tverrsnittenhetene. Dette trekker i en retning av mindre presise estimatorer, og vi gjør oss helt avhengige av en tilstrekkelig variasjon innad i tverrsnittenhetene. Samtidig mister vi også muligheten til å estimere effekten av tidsinvariante variabler.

## 6 Modellspesifikasjoner

### 6.1 Grunnmodellen

For å identifisere inntektselastisiteten til pantelotteriet, konstruerer vi paneldatasettet diskutert i Kapittel 4. Dette inneholder årlig informasjon om pantebeløp og loddsalg fra tilsammen 2851 panteautomater. Denne informasjonen blir koblet sammen med inntekt og andre demografiske faktorer i de 315 kommunene hvor panteautomatene er plassert.

Basert på det tilgjengelige datamaterialet, samt funn og praksis i forskningslitteraturen i tillegg til diskusjonen om datamaterialets begrensinger i Kapittel 4 og 5, bruker vi spillandelen som avhengig variabel og kommunenes brutto medianinntekt som interessevariabel. Videre inkluderer vi mål på sysselsetting, utdanning, alder og befolkningstetthet i den aktuelle kommunen som kontrollvariable, i tillegg til informasjon om hvilken dagligvarekjede som drifter hver enkelt automat. Dette vil gjøre oss i stand til å kontrollere inntektselastisiteten for andre demografiske forhold som også påvirker spilltilbøyeligheten. Denne grunnmodellen muliggjør også ulike empiriske estimeringsmetoder og legger til rette for eventuelle utvidelser og robusthetstester.

$$\begin{aligned} \ln(spillandel)_{it} = & \beta_0 + \beta_1 \ln(inntekt)_{it} + \beta_2 \ln(sysselsetting)_{it} + \beta_3 \ln(18 - 49\text{år})_{it} + \beta_4 \ln(50 - 66\text{år})_{it} \\ & + \beta_5 \ln(66 + \text{år})_{it} + \beta_6 \ln(ingenH.utdanning)_{it} + \beta_7 \ln(KortH.utdannig)_{it} \\ & + \beta_8 \ln(Bostedstype)_{it} + \delta_1 (Dagligvarekjede)_{it} + \kappa_1 (\text{År})_t + \alpha_i + \epsilon_{it} \end{aligned} \quad (16)$$

Her er  $\ln(spillandel)_{it}$  logaritmen til andelen av panteautomatens totale pantebeløp som blir brukt på pantelotteriet for panteautomat  $i$ , i år  $t$ , definert som  $Spillandel = \frac{\text{NOK brukt på pantelotteriet}}{\text{Total pantebeløp}}$ . Videre er  $\ln(inntekt)_{it}$  logaritmen til brutto medianinntekt,  $\ln(sysselsetting)_{it}$  logaritmen til andel sysselsatte over 17 år og  $\ln(Bostedstype)_{it}$  logaritmen til andel bosatt i et tettbebygde strøk for kommunene hvor panteautomatene,  $i$ , er plassert ved år  $t$ . Variablene  $\ln(18 - 49\text{år})$ ,  $\ln(50 - 66\text{år})_{it}$  og  $\ln(66 + \text{år})$  angir logaritmen til andelen av kommunenes befolkning i disse aldersgruppene, mens  $\ln(ingenH.utdanning)_{it}$  og  $\ln(KortH.utdannig)_{it}$  gir oss andelen av kommunens befolkning uten høyere utdanning og andelen med kort høyere utdanning<sup>22</sup>. Både alders- og utdanningsvariablene er gitt for panteautomat  $i$ , i år  $t$ .

<sup>22</sup> Kort høyere utdannede er alle som har tatt mellom 120 og 240 studiepoeng eller har fullført en utdanning i

$(Dagligvarekjede)_{it}$  er en vektor av dummy-variable som gir oss hvilken dagligvarekjede som eier butikken panteautomaten er plassert i.<sup>23</sup>

$(År)_t$  er en vektor av års-dummyer, som blir inkludert for å fange opp effekten av makrovariable som blir utelatt fra grunnmodellen. Dette er spesielt aktuelt for pantelotteriet som har en høy vekst i popularitet gjennom utvalgsperioden, fra 7.4% av totalt pantebeløp i 2012 til 9.4% av totalt pantebeløp i 2015. Dette gir oss en stor endring i den avhengige variabelen over tid, mens endringen i de inkluderte forklaringsvariablene er langt mer moderate. Ved å inkludere årsummyer i grunnmodellen, kan vi kontrollere effektene i forklaringsvariablene for makrovariable som bevissthet om pantelotteriet og en økende popularitet i befolkningen. Med dette unngår vi at den økende populariteten over tid blir tilskrevet de inkluderte forklaringsvariablene.

For å identifisere inntektselastisiteten er både  $\ln(spillandel)_{it}$  og  $\ln(inntekt)_{it}$  gitt som naturlige logaritmer. Dette gir en intuitiv tolkning av de estimerte parameterne, som vil gi oss den prosentvise endringen i spillandelen når medianinntekten øker med 1%. I tillegg til å gi oss elastisiteten mellom den avhengige variabelen og forklaringsvariablene, har en logaritmisk transformasjon flere ønskelige egenskaper. Dersom forholdet mellom spillandelen og en av forklaringsvariablene er non-lineært, vil bruk av log-transformerte variabler ta hensyn til dette, samtidig som Gauss-Markov-forutsetningen om linearitet i parametrene vil holde. Wooldridge (2013, s.185) trekker frem at variabler som er strengt positive gjerne er heteroskedastiske, noe som gjør det problematisk å estimere varians. Ved å log-transformere variablene, vil dette problemet bli begrenset da log-transformasjonen vil redusere spennvidden i variablene. Dette gjør også at estimatet blir mindre sensitivt for store avvik. Utenom dummy-variablene er samtlige av variablene oppgitt som naturlige logaritmer.

Hvilke forklaringsvariable som er inkludert i grunnmodellen bestemmes gjennom en avveining av kontrollvariablenes evne til å frembringe forventningsrette estimater, mot deres tendens til å gjøre estimatene mindre presise. Ved å utelate relevante forklaringsvariable får vi en underspesifisert modell med svekket forklaringskraft og fare for utelatt variabelskjevhet. Dette kan unngås ved å inkludere alle relevante faktorer som påvirker den avhengige variabelen. Samtidig vil et økt antall forklaringsvariable redusere presisjonen ved å øke antall frihetsgrader og forekomst av multikollinearitet. Forklaringsvariablene i denne oppga-

---

universitets- og høyskolesystemet som er kortere enn fem år. (SSB, 2017)

<sup>23</sup>Variabelen for dagligvarekjede er stort sett konstant over tid for alle panteautomater,  $i$ , men blir endret for Rimi- og Ica-butikker da disse ble overtatt av Coop i 2015. Vi tillater derfor  $(Dagligvarekjede)_{it}$  å variere over tid.

ven har blitt valgt ut fra hvilke faktorer som har vært bestemmende og signifikante i liknende studier, samt en vurdering av tilgjengelig datamateriale og empiriske utfordringer diskutert i Kapittel 4 og 5.

## 6.2 Clustering

Som diskutert i Kapittel 5, vil seriekorrelasjon i det enhetsspesifikke restleddet,  $\alpha_i$ , forårsake misledende standardavvik, t-verdier og p-verdier ved bruk av Pooled OLS. Med andre ord vil estimatene kunne fremstå som mer presise enn det er statistisk grunnlag for (Cameron & Miller 2015, s.4).

En vanlig måte å takle dette problemet på er bruk av *cluster-robuste* standardavvik. Her definerer man tverrsnittene som *cluster* av observasjoner som varierer over tid, og tillater heteroskedastisitet og seriekorrelasjon innad i clusterene (Wooldridge 2013, s.688). Dette vil ikke endre de estimerte koeffisientene, men vil føre til mer korrekte estimerte standardavvik, som gir oss et bedre grunnlag for å foreta en troverdig hypotesetesting.

Med unntak av dagligvarekjede- og års-dummyer, er samtlige av denne oppgavens forklaringsvariable gitt på kommunenivå. Dette gjør at panteautomater plassert i samme kommune, vil ha identiske nivåer av disse forklaringsvariablene. Dette kan føre til at restleddene for panteautomater innad i disse kommunene blir korrelert. Etersom de inkluderte forklaringsvariable er gitt på kommunenivå, bør også clusterene bli definert på kommunenivå. Samtlige av denne oppgavens tabeller vil rapportere clusterkorrigerede standardavvik. Cameron & Miller (2015, s.29) anbefaler å bruke minst 50 cluster ved bruk av paneldata, noe som ikke er problematisk da antall kommuner er langt høyere enn dette.

## 6.3 Test av restriksjon

Ved å bruke spillandel som avhengig variabel, antar vi at spillandelen er konstant og uavhengig av det totale pantebeløpet til panteautomatene. For å benytte spillandelen som denne modellens avhengige variabel, er vi derfor prisgitt at dette er en gyldig forenkling av en modell hvor vi har antall kjøpte lodd avhengig variabel, og panteautomatenes totale pantebeløpet som forklaringsvariabel. Ved å betrakte denne modellen, kan vi empirisk teste om

dette er en gyldig forenkling ved å teste restriksjonen,  $\beta_1 = 1$  i likning (17).

$$\ln(\text{SolgteLodd})_{it} = \beta_0 + \beta_1 \ln(\text{Pantebel\o p})_{it} + \kappa_1 X + \alpha_i + \epsilon_{it} \quad (17)$$

Her er  $\ln(\text{SolgteLodd})_{it}$  logaritmen til antall lodd solgt av panteautomatene, mens  $\ln(\text{Pantebel\o p})_{it}$  er panteautomatens totale pantebel\o p, gitt i for panteautomat  $i$  i \a r  $t$ .  $X$  er en vektor av de andre forklaringsvariablene inkludert i grunnmodellen. For \a sikre oss at det estimerte forholdet mellom solgte lodd og totalt pantebel\o p er robust over ulike estimeringsmetoder og inkluderte forklaringsvariable, estimerer vi forholdet med og uten kontrollvariabelvektoren,  $X$ , og tar i bruk b\aa de Fixed Effects og Random Effects som estimeringsmetoder. Dette gir f\o lgende resultater.

Tabell 3: Test av restriksjon

	1)Random Effects	2)Random Effects	3)Fixed Effects	4)Fixed Effects	5)Pooled OLS	6)Pooled OLS
$\ln(\text{Solgte lodd})$						
$\ln\_Pantebel\o p$	1.002 (.0059)	0.996 (.0047)	0.995 (.0068)	0.994 (.0047)	1.02 (.0078)	1.005 (.0066)
Inkludert vektor $X$	Nei	Ja	Nei	Ja	Nei	Ja
$N$	8330	8295	8330	8295	3330	8295

Clusterrobuste Standardavvik i parentes

Dersom forholdet mellom spillandel og pantebel\o p er konstant, vil elastisiteten mellom *Solgte lodd* og *Pantebel\o p* v\aa re lik 1. Modellspesifikasjon 1)-6) viser at den estimerte elastisiteten mellom antall lodd og totalt pantebel\o p har en elastisitet sv\aa rt n\aa re 1. Vi tester om elastisiteten er ulik 1 ved \a sette et signifikansniv\aa lik 0.05 og nullhypotesen;

$$H_0 : \beta_1 = 1 \quad , \quad H_1 : \beta_1 \neq 1$$

Ved \a gjennomf\o re en tosidig t-test p\aa modellspesifikasjon 1), 2), 3), 4), 5) og 6) f\aa r vi henholdsvis p-verdiene,  $p=0.70$ ,  $p=0.43$ ,  $p=0.51$ ,  $p=0.19$ ,  $p=0.001$  og  $p=0.43$ . Dette g\o r at vi med et signifikansniv\aa p\aa 5% ikke kan forkaste at elastisiteten mellom solgte lodd og pantebel\o p er lik 1, i alle modellspesifikasjoner utenom 5). Modellspesifikasjon 5) viser at vi vil f\aa en elastisitet signifikant h\o yere enn 1, ved bruk av Pooled OLS. Dette resultatet b\o r vi v\aa re forsiktige med \a vektlegge, da POLS trolig gir oss feilaktig presise estimater som f\o lge av seriekorrelasjon i det enhetsspesifikke restleddet. Dette ble diskutert i Kapittel 5. Videre i denne oppgaven vil jeg derfor anta at Spillandelen er konstant uansett panteautomatenes pantebe-

løp, og at bruken av  $Spillandel = \frac{\text{NOK brukt på pantelotteriet}}{\text{Total pantebeløp}}$  som avhengig variabel er en gyldig forenkling.





## 7 Resultater fra estimering av økonometrisk modell

I dette kapittelet presenteres resultatene fra økonometriske modellen for inntektselastisiteten. Målet er å undersøke fordelingsvirkningene av pantelotteriet gjennom estimering av inntektselastisiteten ved bruk av økonometrisk modellering. Alt empirisk arbeid er gjennomført i analyseverktøyet *Stata*.

### 7.1 Inntektselastisitet

Inntektselastisiteten til pantelotteriet blir identifisert ved bruk av grunnmodellen i Likning (18) og datamaterialet presentert i Kapittel 4. Jeg vil først presentere resultatene fra ulike estimeringsmetoder. Her vil det også bli diskutert hvilke resultater det bør festes lit til. Etter resultatene fra grunnmodellen er presentert, vil jeg undersøke om inntektselastisiteten og skatteoverveltningen endres over tid, i tillegg til å se om resultatene er robuste for andre mål på inntekt. Gjennom kapittelet vil jeg sammenlikne denne oppgavens resultater med resultater fra tidligere forskning.

$$\begin{aligned} \ln(\text{spillandel})_{it} = & \beta_0 + \beta_1 \ln(\text{inntekt})_{it} + \beta_2 \ln(\text{sysselsetting})_{it} + \beta_3 \ln(18 - 49\text{år})_{it} + \beta_4 \ln(50 - 66\text{år})_{it} \\ & + \beta_5 \ln(66 + \text{år})_{it} + \beta_6 \ln(\text{ingenH.utdanning})_{it} + \beta_7 \ln(\text{KortH.utdannig})_{it} \\ & + \beta_8 \ln(\text{Bostedstype})_{it} + \delta_1 (\text{Dagligvarekjede})_{it} + \kappa_1 (\text{År})_t + \text{Restledd} \end{aligned} \quad (18)$$

Alle estimeringsmetodene tar utgangspunkt i grunnmodellen i Likning (18). Resultater for tidsdummyene og dagligvarekjede-dummyene vil ikke bli inkludert i tabellene i dette kapittelet, men ligger vedlagt i appendikset.

#### 7.1.1 Pooled OLS

Tabell 4 viser resultatene fra *Pooled OLS* (POLS)-estimeringen av grunnmodellen. Kolonne (1)-(6) viser resultatene fra modellspesifikasjoner med stigende antall inkluderte kontrollvariable, hvor Kolonne (1) kun inkluderer interessevariabelen, inntekt, og årsummyer mens Kolonne (6) inkluderer samtlige kontrollvariable.

Ved POLS-estimering er samtlige inntektselastisiteter estimert høyere enn 2, og ligger i spennet 2.06-3.63. Brutto inntektselastisitet estimert i Kolonne (1), ligger på 2.41, og faller til

2.01 når vi inkluderer alle kontrollvariable i Kolonne (6).

Som diskutert i Kapittel 5, er POLS estimeringer avhengige av strengt eksogene forklaringsvariable for å få konsistente og forventningsrette parameterestimatorer,  $E(\alpha_i|x_i) = 0$ . Dersom denne antagelsen holder, vil POLS nyttiggjøre seg av variasjon både innad og mellom tversnittenhetene, og vil dermed kunne gi presise estimater. Som vi ser i Tabell 4 har samtlige av de estimerte inntektselastisitetene en t-verdi høyere enn 10, og later til å være svært presise. Tross tilsynelatende presist estimerte inntektselastisiteter og sterk signifikans blant parameterestimaterne, kan vi ikke anta at disse resultatene er pålitelige. Dette kommer av at POLS vil være heftet med et seriekorrelert restledd som følge av at det enhetsspesifikke restleddkomponenten,  $\alpha_i$ , vil være konstant for alle tidsperioder. Dette vil gi oss kunstig lave standardavvik, som igjen gir oss et galt inntrykk av POLS-estimatorenes presisjon. t-verdiene og signifikansen vi ser i Tabell 4 vil dermed være ugyldige. Dette gjør at vi må ta i bruk alternative estimeringsmetoder, som *Random Effects* eller *Fixed Effects*

Tabell 4: Pooled OLS

	(1)	(2)	(3)	(4)	(5)	(6)
	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)
ln(Inntekt)	2.415*** (29.25)	3.633*** (31.02)	3.197*** (25.26)	2.106*** (11.07)	2.066*** (10.13)	2.016*** (10.17)
ln(Sysselsetting)		-2.862*** (-15.74)	-2.443*** (-12.39)	-2.084*** (-10.70)	-2.031*** (-8.96)	-1.728*** (-7.82)
ln(18-49år)			1.843*** (9.65)	0.782*** (3.37)	0.774*** (3.31)	0.881*** (3.86)
ln(50-66år)			0.109 (0.61)	0.00538 (0.03)	0.0200 (0.11)	0.110 (0.63)
ln(66+år)			0.607*** (7.31)	0.264** (2.86)	0.264** (2.85)	0.256** (2.87)
ln(Ingen H.utdanning)g				-0.780*** (-4.94)	-0.802*** (-4.62)	-0.762*** (-4.48)
ln(Kort H.utdanning)				-0.0206 (-0.24)	-0.0387 (-0.37)	-0.0860 (-0.84)
ln(Tettsted)					0.0179 (0.45)	-0.00350 (-0.09)
Tidsdummyer	Ja	Ja	Ja	Ja	Ja	Ja
Dagligvarekjede-dummyer	Nei	Nei	Nei	Nei	Nei	Ja
Konstantledd	-33.47*** (-31.82)	-36.90*** (-34.48)	-30.17*** (-23.21)	-19.86*** (-10.53)	-19.58*** (-10.19)	-20.18*** (-10.82)
N	8330	8317	8317	8317	8295	8295
R <sup>2</sup>	0.24	0.30	0.35	0.36	0.36	0.40

t-verdier basert på clusterrobuste standardavvik i parentes.

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

### 7.1.2 Random Effects

Random Effects (RE) unngår seriekorrelerte restledd og inkonsistente standardavvik ved å bruke GLS-estimerer som transformerer bort de seriekorrelerte restleddene. Som diskutert i Kapittel 5, vil Random Effects, under antagelsen om strengt eksogene forklaringsvariable, utnytte variasjon i både tids- og tverrsnitts-dimensjonen og gi oss konsistente og effisiente estimatorer. Resultatene fra RE-estimeringen presenteres i Tabell 5.

Samtlige RE-estimeringer viser en inntektselastisitet som er høyere enn 1 og signifikant ulik 0. De ulike modellspesifikasjonene viser at inntektselastisiteten er sensitiv til hvilke kontrollvariable vi inkluderer. Variasjonen til inntektselastisiteten er påfallende stor, og varierer fra 1.1 til 2.7, avhengig av antall kontrollvariable. Inntektselastisiteten er på sitt høyeste ved 2.7, (Kolonne (2)), når det kun kontrolleres for sysselsetting, mens den er på sitt laveste i Kolonne (5), hvor inntektselastisiteten er estimert til 1.05. Det ser ut til at den estimerte inntektselastisiteten faller med antall inkluderte kontrollvariable, hvilket indikerer at inkluderingen av disse variablene har en kontrollerende effekt.

RE resultatene later til å følge samme mønster som ved bruk av POLS, med en høy brutto inntektselastisitet i modellspesifikasjon (1), som reduseres med antall inkluderte kontrollvariable. Samtidig er det tydelig at GLS-transformasjonen har moderert den estimerte inntektselastisiteten og redusert både signifikans og presisjon i modellen. t-verdiene er nå redusert fra en verdi på over 10, og ligger nå i spennet 3.3-8.2. Dette indikerer at POLS-estimatene var heftet med en problematisk seriekorrelasjon og inkonsistente standardavvik.

En inntektselastisitet over 1, tilsier at en økning av inntekten på 10% vil gi en økning i spillandelen på over 10%. Spillkonsumet øker altså proporsjonalt mer enn inntekten, og vil med det utgjøre en større del av konsumentens budsjett, når inntekten øker. Ved å se på lotterispill som en implisitt skatt, kan vi argumentere for at pantelotteriets skattebyrde er progressivt økende med inntektsstørrelse om inntektselastisiteten er høyere enn 1. Test av modellspesifikasjonene i Kolonne (1) til (6) i Tabell 5, med et signifikansnivå på 5% og en nullhypotese satt:

$$H_0 : \beta_1 = 1 \quad , \quad H_1 : \beta_1 \neq 1$$

viser at vi kan forkaste nullhypotesen om inntektselastisitet lik 1 i modellspesifikasjon (1), (2) og (3), med p-verdier på henholdsvis  $p=0.003$ ,  $p=0.000$  og  $p=0.000$ . Vi kan derimot ikke forkaste  $H_0$  i modellspesifikasjon (4), (5) eller (6). Dette gjør at vi ikke har et statistisk grunnlag

Tabell 5: Random Effects

	(1)	(2)	(3)	(4)	(5)	(6)
	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)
ln(Inntekt)	2.105*** (5.65)	2.762*** (6.62)	2.436*** (8.12)	1.141*** (3.64)	1.058*** (3.33)	1.113*** (3.55)
ln(Sysselling)		-1.517*** (-4.27)	-1.270*** (-4.18)	-0.959*** (-3.32)	-0.819** (-2.67)	-0.754* (-2.55)
ln(18-49år)			1.672*** (4.00)	0.286 (0.63)	0.223 (0.47)	0.284 (0.61)
ln(50-66år)			-0.171 (-0.56)	-0.383 (-1.27)	-0.387 (-1.27)	-0.372 (-1.24)
ln(66+år)			0.680*** (3.79)	0.353 (1.87)	0.363 (1.89)	0.386* (2.03)
ln(Ingen H.utdanning)				-0.537 (-1.92)	-0.593* (-2.10)	-0.557* (-1.98)
ln(Kort H.utdanning)				0.428* (2.36)	0.373* (2.00)	0.343 (1.84)
ln(Tettsted)					0.0546 (1.08)	0.0299 (0.62)
Tidsdummyer	Ja	Ja	Ja	Ja	Ja	Ja
Dagligvarekjede-dummyer	Nei	Nei	Nei	Nei	Nei	Ja
Konstantledd	-29.54*** (-6.26)	-31.50*** (-7.13)	-25.89*** (-7.55)	-12.40*** (-3.56)	-12.05*** (-3.46)	-12.96*** (-3.76)
<i>N</i>	8330	8317	8317	8317	8295	8295
<i>R</i> <sup>2</sup>	0.23	0.29	0.33	0.34	0.34	0.38

*t*-verdier basert på clusterrobuste standardavvik i parentes.

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

for å si at pantelotteriet er progressivt beskattende når vi kontrollerer for utdanningsnivå. Resultatene vi får ved bruk av Random Effects som estimeringsmetode antyder at pantelotteriet er progressivt beskattende, men at spilltilbøyeligheten i ulike inntektsgrupper kan forklare av andre demografiske forhold i tillegg til inntekten. Disse resultatene skiller seg ut fra estimeringer av inntektselastisiteten i sammenlignbare studier, som stort sett finner en signifikant regressivitet i lotteriers implisitte beskatning. Perez & Humphreys (2013, s.930) presenterer i sitt litteratursammendrag inntektselastisiteten til 17 studier av lotteriers fordelings effekter. Kun et av studiene, Mikesell (1989), har en inntektselastisitet høyere enn 1. De resterende 16 studiene har en inntektselastisitet i spennet -1.1 til 0.92.

Selv om den estimerte inntektselastisiteten i Tabell 5 ikke sammenfaller med inntektselastisitetene ellers i litteraturen, underbygger en inntektselastisitet over 1 de progressive resultatene vi fikk i estimeringen av Suits-indeksene i Kapittel 4.

Både Pooled OLS og Random Effects som estimeringsmetoder avhenger av standardan-

tagelsen om fravær av korrelasjon mellom det enhetsspesifikke restleddet og forklaringsvariablene,  $E(\alpha_i|x_i) = 0$ . Dette betyr at ingen av de inkluderte forklaringsvariablene kan være korrelert med andre butikkspesifikke faktorer som også kan påvirke spilltilbøyeligheten. Endogenitetsproblemer og muligheten for brudd på antagelsen om strengt eksogene forklaringsvariable er diskutert i Kapittel 5, hvor det framkommer at denne antagelsen høyst sannsynlig vil bli brutt som en følge av butikkspesifikke utelatte variable som er korrelert med både den avhengige variabelen og forklaringsvariablene. Denne mistanken blir bekreftet ved å observere hvor sensitiv inntektselastisiteten er til forklaringsvariablene som er inkludert i grunnmodellen. Vi bør derfor anta at både Pooled OLS og Random Effects estimatene er heftet med en utelatt variabel-skjevhet.

### 7.1.3 Fixed Effects

Inkonsistente estimater som følge av brudd på antagelsen av strengt eksogene forklaringsvariable tilsier at vi bør ta i bruk estimeringsmetoder som ikke er avhengige av at denne antagelsen holder. Estimeringsmetoden *Fixed Effects*, FE, frir oss fra denne antagelsen, og gjør oss i stand til å få konsistente estimater på tross av korrelasjon mellom det enhetsspesifikke restleddet og forklaringsvariablene. Ved å innføre dummyvariabler for hver tverrsnittenhets, eller utføre en within-transformasjon hvor vi subtraherer alle elementene med sine respektive enhetsgjennomsnitt, vil vi fjerne alle enhetsfaste effekter, inkludert det problematiske enhetsspesifikke restleddet,  $\alpha_i$ .

Tabell 6 viser oss resultatene for grunnmodellen ved bruk av Fixed Effects som estimeringsmetode. Vi ser raskt at det kun er *Kort H.utdanning* som har en signifikant effekt, mens resten av forklaringsvariablene ikke har noen signifikante effekter. Videre viser Tabell 6 at den estimerte inntektselastisiteten er langt lavere ved bruk av Fixed Effects enn ved bruk av Pooled OLS og Random Effects. Dette er ingen overraskelse da RE estimatene legger seg mellom POLS- og FE-estimatene, avhengig av hvor stor del den enhetsspesifikke restleddvariansen utgjør av den totale restleddvariansen. Med kun fire tidsperioder, vil seriekorrelerte enhetsspesifikke restledd representere et reelt, men begrenset problem. Dette kan forklare at RE estimatene ligger noe nærmere POLS estimatene enn FE estimatene.

Det fremste problemet med FE som estimeringsmetode for denne oppgavens datasett, er at within-transformasjonen rensker bort all variasjon mellom tverrsnittenheter (between-variasjon). Dette er spesielt problematisk for dette paneldatasettet da between-variasjonen

Tabell 6: Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)
	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)
ln(Inntekt)	0.304 (0.52)	0.271 (0.39)	0.201 (0.31)	-0.922 (-1.66)	-0.882 (-1.56)	-0.894 (-1.56)
ln(Sysselsetting)		0.0824 (0.14)	0.273 (0.59)	0.309 (0.93)	0.339 (1.01)	0.364 (1.08)
ln(18-49år)			0.437 (0.44)	-0.853 (-1.26)	-0.766 (-1.11)	-0.736 (-1.08)
ln(50-66år)			-0.870 (-1.83)	-0.634 (-1.62)	-0.648 (-1.64)	-0.650 (-1.64)
ln(66+år)			0.772 (1.90)	0.544 (1.63)	0.570 (1.72)	0.573 (1.70)
ln(Ingen H.utdanning)				-0.351 (-0.60)	-0.411 (-0.70)	-0.193 (-0.28)
ln(Kort H.utdanning)				1.341*** (4.91)	1.398*** (5.00)	1.426*** (5.18)
ln(Tettsted)					-0.150 (-1.90)	-0.170* (-2.09)
Tidsdummyer	Ja	Ja	Ja	Ja	Ja	Ja
Dagligvarekjede-dummyer	Nei	Nei	Nei	Nei	Nei	Ja
Konstantledd	-6.584 (-0.88)	-6.507 (-0.85)	-5.856 (-0.88)	9.053 (1.36)	8.548 (1.26)	8.671 (1.27)
<i>N</i>	8330	8317	8317	8317	8295	8295
<i>R</i> <sup>2</sup>	0.454	0.455	0.464	0.479	0.480	0.484

*t*-verdier basert på clusterrobuste standardavvik i parentes.

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

utgjør en svært stor andel av datasettets totale variasjon. Tabell 12 i appendikset viser hvordan den totale variasjonen fordeles mellom between- og within-variasjon. Her ser vi at forklaringsvariablene som beskriver kommunenes alderssammensetning, utdanningsnivå eller befolkningstetthet varierer svært lite over tid. Dette skyldes at datasettet kun strekker seg over fire år, og at denne typen demografiske variable varierer lite over tid.

Problemet med manglende variasjon innad i tverrsnittene er ikke unikt for denne oppgaven. Av de få studiene av inntektselastisiteten for lotterier som bruker paneldata, har jeg ikke lyktes i å finne noen som tar i bruk FE. Baker et al. (2016, s.132) bruker områdedata for 52 spanske provinser over 522 uker. De konkluderer med at en Fixed Effects-modell ikke er gjennomførbar grunnet for liten within-variasjon i modellens forklaringsvariable. Resultatene i Tabell 6 tyder på at en kort utvalgsperiode og demografiske forklaringsvariable på kommunenivå gjør at en within-transformasjon fjerner problematisk mye av datasettets variasjon. Fixed Effects er som estimeringsmetode helt avhengig av tilstrekkelig variasjon innad

i tverrsnittene. Ved å transformere bort all variasjon mellom tverrsnittene, mister vi også alle muligheter for å påvise signifikans. Som studiet til Baker et al. (2016), er ikke betingelsene for en grundig Fixed Effects estimering tilstede i denne oppgavens datasett. Dette gjør at vi videre i oppgaven vil gjøre som Baker et al. (2016), og benytte Random Effects som estimeringsmetode. Vi må i tolkningen av RE-resultatene i Tabell 5 ta forbehold om at disse resultatene kan være heftet med utelatt variabelskjevhet ettersom disse avhenger av antagelsen om strengt eksogene forklaringsvariable.

#### **7.1.4 Effekt av øvrige forklaringsvariable.**

Som ved estimeringen av inntektselastisiteten vil resultatene fra Random Effects-estimeringen bli vektlagt i presentasjonen av effekten til de øvrige forklaringsvariablene. Resultatene fra POLS- og FE-estimeringer kan finnes i Tabell 4 og 6, og Tabell 13 og 15 i appendikset.

I likhet med inntektselastisiteten, skiller også resultatene fra forklaringsvariablene seg fra sammenlignbare studier. De tre inkluderte aldersvariablene, *18-49år*, *50-66år* og *66+år*, indikerer et U-formet forhold mellom spilltilbøyelighet og alder, hvor andelen av befolkningen i aldersgruppene 18-49år og 66+år trekker spillandelen opp, mens andelen i aldersgruppen 50-66år trekker spillandelen ned. Dette er motsatt av resultatene til Beckert & Lutter (2008, s.483), som finner at forholdet mellom lotterikonsum og alder best blir beskrevet ved en invertert U-form. Effekten av alderssammensetningen i kommunene er forøvrig kun signifikant i modellspesifikasjon (3) og (6), og bør derfor ikke tillegges mye vekt.

Andelen i kommunene uten høyere utdanning har en signifikant negativ elastisitet på -0.59 og -0.56 i henholdsvis modellspesifikasjon (5) og (6), mens andelen med kort høyere utdanning har en signifikant positiv elastisitet på 0.43 og 0.37 i modellspesifikasjon (4) og (5). Dette indikerer at utdanningsnivået har en positiv effekt på befolkningens tilbøyelighet til å spille på pantelotteriet. Hansen (1995, s.393) bruker en lignende definisjon av utdanningsnivå i sitt studie, og finner også at andelen med 16 år eller lengre utdanning har en positiv effekt på et fylkes loddsalg.

Viktigst for denne oppgaven er utdannings-variablenes kontrollerende effekt på inntektselastisiteten. Vi inkluderer disse for å unngå en utelatt variabelskjevhet (OVB) som oppstår dersom utdanningsnivået både påvirker spilltilbøyeligheten og inntektsnivået i en kommune. Ved inkludering av utdanningsvariablene i modellspesifikasjon (4), ser vi at inntektselastisiteten



tisiteten faller fra 2.44 til 1.14, noe som antyder at inntektselastisiteten i modellspesifikasjon (1) og (2) er heftet med en tydelig OVB, og at vi isolerer inntektseffekten langt bedre ved inkluderingen av disse.

*Syssetningen* har signifikant negativ effekt på spilltilbøyeligheten i samtlige av modellspesifikasjonene den er inkludert, og har i Kolonne (6) en elastisitet på -0.75. Dette er ikke ulikt funnene til Mikesell (1994), som finner at deltagelsen i lotterier øker med arbeidsledigheten. I motsetning til de andre kontrollvariablene, viser *syssetningen* en klar signifikant effekt over alle modellspesifikasjoner i Tabell 5.

Kontrollvariabelen *Tettsted* har ingen signifikante effekter. Dette skiller seg fra resultatene til Hansen (1995, s.394) og Heavey (1978, s.424) som finner at loddsalget per capita øker med befolkningstetthet i urbane strøk. At denne oppgavens resultater av bostedstype skiller seg fra amerikanske resultater er ingen overraskelse, da norske byer og tettsteder minner lite om amerikanske byer med rike forsteder og sosialt vanskeligstilte bykjerne.

Tabell 14 i appendikset viser at dagligvarekjedene Rimi, Ica, Centra og Meny har en signifikant positiv effekt på spillandelen med henholdsvis 0.07, 0.07, 0.15 og 0.13. At Centra og Meny som breddesortiment-kjeder har den klart høyeste effekten på spillandelen føyer seg inn i et mønster som skisserer pantelotteriet som et progressivt beskattende lotteri. Noe som svekker denne påstanden er at breddesortiments-kjeden Spar er eneste dagligvarekjede med en signifikant negativ effekt på spillandelen med -0.14. Butikkeffekter ved bruk av POLS og FE finnes også i appendikset i Tabell 13 og 15.

## 7.2 Tolkning av inntektselastisiteten

Tabell 5 viser at inntektselastisiteten går fra 2.1 når vi kun inkluderer inntekt som forklaringsvariabel, til 1.1, når alle forklaringsvariable i grunnmodellen er inkludert. Modellspesifikasjon (6) viser at den kontrollerte effekten av inntekt på spilltilbøyeligheten ikke er signifikant høyere enn 1, og at vi dermed ikke kan forkaste hypotesen om at pantelotteriet er regressivt eller proporsjonalt beskattende. Dersom vi ønsker å isolere inntektseffekten på spilltilbøyelighet vil modellspesifikasjon (6) være best egnet, da denne kontrollerer inntektseffekten mest. Fokuset i denne oppgaven er derimot ikke å isolere inntektselastisiteten best, men å se

hvordan pantelotteriet som en implisitt skatt belaster ulike inntektsgrupper. Om inntekten til pantelotteriets kunder kommer av ulikt utdanningsnivå, ulik alderssammensetning eller ulikt sysselsettingsnivå er av sekundær interesse. Baker et al. (2016, s.136) argumenterer for at man kun skal inkludere inntekten som eneste forklaringsvariabel for å identifisere lotteri-ers faktiske fordelings effekter. De mener dersom man ser på inntektselastisiteten, *ceteris paribus*, vil man ende opp med et utvannet inntrykk av den egentlige skattefordelingen. Etter som denne oppgaven også inkluderer Suits-indeksers som et rent deskriptivt mål på pantelotteriets fordelings effekter, er det ikke nødvendig å følge eksempelet til Baker et al. (2016) for å få enda et deskriptivt progressivitetsmål. Kombinasjonen av både Suits-indeksers og inntektselastisitet kontrollert for andre demografiske effekter burde dermed gjøre oss godt rustet til å si noe om pantelotteriets faktiske fordelings effekter, og kildene til dette.

## 7.3 Robusthet og utvidelser

### 7.3.1 Robusthet over tid

Som et ganske nytt spillkonsept, bør vi ikke anta at lotteriets kundemasse holder seg konstant over tid, men at både kunder og spillvaner endres etterhvert som pantelotteriet blir mer utbredt. Dette reflekteres av spillandelens vekst fra 3,7% i 2009 til 10% i 2017. Med en slik endring i spillandel og omsetning, er det ønskelig å se om også skatteprogressiviteten endres gjennom utvalgsperioden. At Suits-indeksene i Tabell 1 viser en klar bevegelse mot en mer proporsjonal beskatning, indikerer også at vi ikke kan anta at inntektselastisiteten er konstant gjennom utvalgsperioden.

Få studier av lotterier benytter paneldata, og det er dermed gjort lite forskning på hvordan lotteriets fordelings effekter endres over lotterienes livsløp. Av studiene som er gjennomført, finner Miyazaki et al. (1998, s.164) at lotterier blir signifikant mindre regressive gjennom lotteriets livsløp. Ved å se på inntektselastisitet for de ulike utvalgsårene, kan vi se om dette også er tilfellet for pantelotteriet.

Den estimerte grunnmodellen i Tabell 5 forutsetter at effekten av forklaringsvariablene er konstante over tid. For å se på hvordan inntektselastisiteten endres over tid, tar vi utgangspunkt i denne modellen, og inkluderer interaksjonsleddene *Inntekt \* 2013*, *Inntekt \* 2014* og *Inntekt \* 2015*<sup>24</sup>. Dette vil fange opp eventuell ikke-linearitet i inntektselastisiteten og gjør oss i stand til å se om inntektselastisiteten endres gjennom utvalgsperioden.

---

<sup>24</sup>2012 blir brukt som referanseår

Tabell 7: Random Effects med Interaksjonsledd

	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)
ln(Inntekt)	2.613*** (7.52)	3.219*** (9.06)	2.852*** (9.53)	1.513*** (4.75)	1.336*** (4.22)	1.388*** (4.42)
ln(Inntekt*2013)	-0.577*** (-6.96)	-0.574*** (-6.64)	-0.574*** (-6.46)	-0.581*** (-6.90)	-0.593*** (-6.89)	-0.589*** (-6.79)
ln(Inntekt*2014)	-0.753*** (-6.00)	-0.746*** (-5.56)	-0.732*** (-5.24)	-0.783*** (-6.09)	-0.817*** (-6.19)	-0.830*** (-6.34)
ln(Inntekt*2015)	-0.725*** (-5.45)	-0.684*** (-4.53)	-0.682*** (-4.41)	-0.853*** (-5.81)	-0.908*** (-6.00)	-0.923*** (-6.20)
Inkluderte variabler fra modellspesifikasjon:	(1)	(2)	(3)	(4)	(5)	(6)
Konstantledd	-36.01*** (-8.18)	-37.78*** (-9.48)	-31.44*** (-8.92)	-18.71*** (-5.22)	-18.06*** (-5.02)	-18.92*** (-5.26)
N	8330	8317	8317	8317	8295	8295
R <sup>2</sup>	0.24	0.29	0.33	0.34	0.35	0.38

*t*-verdier basert på clusterrobuste standardavvik i parentes.

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

Utviklingen i Tabell 7 er tydelig og viser at inntektselastisiteten er fallende og signifikant lavere enn i referanseåret 2012. F-test med nullhypotesen  $\beta_2 = \beta_3 = \beta_4 = 0$  forkastes med en *p*-verdi på 0.0000, og viser at en modellspesifikasjon som tillater endring i inntektselastisiteten over tid er bedre egnet til å beskrive inntektselastisiteten. Dette gjør at grunnmodellen ikke er en gyldig forenkling av en modell med interaksjonsleddene *Inntekt* \* 2013, *Inntekt* \* 2014 og *Inntekt* \* 2015.

For å tillate at effektene til samtlige forklaringsvariable kan endres over tid, estimerer vi regresjoner med tverrsnittdata fra hvert år. Tabell 8 viser resultater fra OLS på modellspesifikasjon (6) for hvert av utvalgsårene. Her kan vi se at den estimerte inntektselastisiteten faller fra 2.05 i 2012 til 1.7 i 2015.

De økonometriske resultatene i Tabell 7 og 8 samsvarer med den fallende Suits-indeksen i Tabell 1. At også inntektselastisiteten reduseres over tid styrker inntrykket av at pantelotteriet blir mindre progressivt beskattende gjennom utvalgsperioden. Dette skiller seg fra funnene til Miyazaki et al. (1998). Det bør samtidig nevnes at fordelingseffekten i begge studier blir mer proporsjonalt beskattende over tid.

### 7.3.2 Robusthet MHT alternativt inntektsmål

For å se om hvorvidt estimeringen av inntektselastisiteten avhenger av hvordan vi definerer inntekten, er det ønskelig å se om resultatene endres om vi endrer definisjonen av inntekt til gjennomsnittsinntekt. Bruk av gjennomsnittlig inntekt kan også gi oss mer presise estimater ettersom denne varierer mer enn medianinntekten. Tabell 9 viser at standardavviket til

Tabell 8: OLS på Tverrsnitt

	2012	2013	2014	2015
	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)
ln(linntekt)	2.054*** (4.24)	1.954*** (4.72)	1.951*** (4.09)	1.704** (3.31)
ln(Syssellsetting)	-1.172* (-2.32)	-1.549*** (-3.34)	-1.958*** (-4.21)	-1.772*** (-3.63)
ln(18-49år)	0.967 (1.76)	0.781 (1.44)	1.308* (2.22)	0.574 (0.99)
ln(50-66år)	0.231 (0.58)	0.129 (0.35)	0.373 (0.78)	0.0694 (0.15)
ln(67+år)	0.0628 (0.27)	0.210 (0.94)	0.341 (1.31)	0.280 (1.17)
ln(IngenH.utdanning)	-0.919* (-2.11)	-0.815* (-2.17)	-0.811 (-1.94)	-0.961* (-2.30)
ln(KortH.Utdanning)	-0.157 (-0.69)	-0.179 (-0.82)	-0.135 (-0.58)	-0.155 (-0.66)
ln(Tettsted)	0.00339 (0.05)	0.0627 (0.83)	-0.0362 (-0.57)	0.0274 (0.35)
Dagligvarekjede-dummyer	Ja	Ja	Ja	Ja
Konstantledd	-23.17*** (-5.04)	-20.29*** (-5.00)	-17.44*** (-3.77)	-16.37** (-3.15)
<i>N</i>	1878	2055	2041	2321
<i>R</i> <sup>2</sup>	0.42	0.39	0.39	0.35

*t*-verdier basert på clusterrobuste standardavvik i parentes.

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

inntektsvariabelen øker fra NOK 27 237 til NOK 49 518 ved bruk av snittinntekt.

Tabell 9: Ulike innteksmål

	Medianinntekt	Gj.snittinntekt
Gjennomsnitt	348 919	403 557
Standardavvik	27 237	49 518
Maks.	420 739	590 798
Min.	244 800	283 600

Tabell 10 viser at samtlige av de estimerte inntektselastisitetene reduseres ved bruk av kommunenes gjennomsnittlige inntekt som inntektsvariabel. I modellspesifikasjon (1) faller inntektselastisiteten fra 2.1 til 1.5, mens den i modellspesifikasjon (2) faller fra 1.1 til 0.97. Samtlige *t*-verdier er høyere ved bruk av gjennomsnittsinntekt, noe som tyder på mer presise estimater. Som ved bruk av medianinntekt, er det også i dette tilfellet kun modellspesifikasjon (1), (2) og (3) som gir en inntektselastisitet signifikant ulik 1. Forklaringskraften,  $R^2$ , er kun marginalt høyere ved enkelte modellspesifikasjoner, ved bruk av gjennomsnittlig inn-

tekt. Det ble også gjennomført FE- og POLS-estimeringer med gjennomsnittsinntekten som inntekts-variabel. Dette ga ingen utslag på de estimerte inntektselastisitetenes signifikans. Disse resultatene finnes i Tabell 17 og 18 i appendikset.

Tabell 10: Gjennomsnittlig inntekt, Random Effects

	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)
ln(Gjennomsnittlig inntekt)	1.489*** (7.19)	1.775*** (6.91)	1.559*** (8.61)	1.002*** (4.19)	0.977*** (4.07)	0.966*** (3.97)
Inkluderte variabler fra modellspesifikasjon:	(1)	(2)	(3)	(4)	(5)	(6)
Konstantledd	-21.88*** (-8.28)	-20.57*** (-8.39)	-16.14*** (-7.63)	-9.876*** (-3.85)	-9.915*** (-3.86)	-10.19*** (-3.90)
<i>N</i>	8330	8317	8317	8317	8295	8295
<i>R</i> <sup>2</sup>	0.28	0.31	0.35	0.35	0.35	0.38

*t*-verdier basert på clusterrobuste standardavvik i parentes.

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

Totalinntrykket er at hverken resultatene til inntektselastisiteten eller Suits-indeksene er problematisk sensitive til hvordan vi definerer inntekt i denne oppgaven.

## 8 Oppsummering og konklusjon

Denne oppgaven har som formål å undersøke fordelingseffektene til pantelotteriet. Dette er et veldedig lotteri hvor eneste mulige innsats er pantepenger man får ved retur av tomgods. Pantelotteriets overskudd går til Røde Kors, og har en årlig omsetningsramme på NOK 300 millioner. Pantelotteriet er ett av fem nye landsdekkende lotterier drevet av humanitære og samfunnsnyttige organisasjoner. Disse lotteriene utgjør tilsammen et lotterimarked med en årlig øvre grense på NOK 1.5 milliarder, og ble i 2017 tildelt konsesjoner på ni år. I tillegg til å være driftet av samfunnsnyttige organisasjoner, må disse lotteriene følge flere særkrav som gjør at de skiller seg fra tradisjonelle lotterier. De kan blant annet ikke være avhengighetsskapende. Som det eneste lotteriet som har blitt driftet under disse vilkårene i flere år, tar denne oppgaven i bruk datamateriale fra pantelotteriet for å undersøke de potensielle fordelingsvirkningene av de nye landsdekkende lotteriene.

Datasettet fra pantelotteriet strekker seg fra 2012 til 2015, og inneholder informasjon fra alle pantetransaksjoner fra 2357 panteautomater med lotterifunksjon. Denne informasjonen ble så koblet med områdedata fra kommunene hvor panteautomatene er plassert, hvor kommunenes brutto medianinntekt blir brukt som mål på inntekt.

Siden 35% av omsetningen betales tilbake som premier, innebærer pantelotteriet i likhet med andre lotterier en implisitt skatt for deltakerne. For å beskrive fordelingsvirkningene til pantelotteriets implisitte skatt, bruker denne oppgaven to ulike mål. Suits-indeks brukes som et deskriptivt mål på hvordan den implisitte skattebyrden til pantelotteriet fordeles blant ulike inntektsgrupper. Pantelotteriets inntektselastisitet brukes for å identifisere en mer isolert effekt av inntekt på spilltilbøyelighet.

I motsetning til et stort flertall av studier på fordelingseffektene til lotterier og pengespill, finner denne oppgaven at pantelotteriet har en positiv Suits-indeks for samtlige år i utvalgsperioden, med en gjennomsnittlig Suits-indeks på 0.053. En positiv Suits-indeks indikerer at en skatt eller avgift er progressivt beskattende. Suits-indeksene til 11 andre studier gjort på pengespill viser at disse er regressivt beskattende med negative Suits-indeks i spennet -0.46 til -0.03 (Perez & Humphreys 2013, s.930).

Ved å estimere en økonometrisk modell basert på det konstruerte paneldatasettet, får vi en estimert inntektselastisitet mellom 2.7 og 1.1, avhengig av antall inkluderte kontrollvariable. Modellspekifikasjonen med inntekt som eneste forklaringsvariable, gir oss en inntekts-

elastisitet lik 2.1 og er signifikant høyere enn 1. Når vi kontrollerer inntektselastisiteten for alle kontrollvariable i datasettet, faller den til 1.1, og er ikke lenger signifikant høyere enn 1. Disse resultatene skiller seg også fra andre studier av inntektselastisitetene til lotterier og pengespill. Perez & Humphreys (2013) viser i sitt litteratursammendrag at 16 av 17 studier finner en inntektselastisitet under 1. Resultatene fra de estimerte inntektselastisitetene underbygger resultatene fra Suits-indeksenene og styrker inntrykket av pantelotteriet som progressivt beskattende. Samtidig viser de økonometriske resultatene at selv om skattebyrden faller på befolkningen med høyere inntekt, er det også andre demografiske faktorer som er drivende for lotterikonsumet, deriblant sysselsetting, utdanningsnivå og alderssammensetning.

Videre utnytter denne oppgaven mulighetene et paneldatasett gir til å se om pantelotteriets fordelingseffekter endres over tid. Vi finner at Suits-indeksen faller fra 0.083 i 2012 til 0.033 i 2015. Ved å inkludere interaksjonsledd i den økonometriske modellen som tillater at inntektselastisiteten endres over tid, finner vi at også inntektselastisiteten er fallende gjennom utvalgsperioden. Det samme gjelder når vi estimerer inntektselastisiteten med tverrsnittdata for hvert enkelt år. Endringen i skatteprogressiviteten er overraskende stor, men dette må sees i lys av at pantelotteriet i perioden 2012-2015 var et ganske nytt lotteri, med en stor vekst i omsetning gjennom utvalgsperioden.

Både Suits-indeksene og inntektselastisiteten ble også estimert med bruk av gjennomsnittsinntekt i kommunen som mål på inntekt. Disse resultatene viser i stor grad like fordelingseffekter og utvikling over tid som ved bruk av medianinntekt.

Etttersom denne oppgavens resultater skiller seg fra sammenliknbare studier, vil ikke det teoretiske rammeverket som har blitt brukt til forklare fordelingseffekten av regressivt beskattende lotterier være like anvendelig til å forklare pantelotteriets fordelingseffekter. Det er heller ikke innenfor denne oppgavens problemstilling å forklare hvorfor skattebyrden fordeles som den gjør. Oppgavens datasett har ingen forutsetninger for å gjøre dette. Ved å se på hvordan pantelotteriet skiller seg fra tradisjonelle lotterier, kan vi drøfte mulige kilder til de spesielle fordelingseffektene av pantelotteriet.

Pantelotteriet skiller seg klart fra andre lotterier ved at lodd kun kan kjøpes ved retur av tomgods. Dette påvirker både rammene og miljøet for det økonomiske valget mellom en sikker lav godtgjørelse eller en usikker potensielt høy premie. Spillavhengige og andre pro-

blemspillere blir også effektivt ekskludert fra lotteriet i og med at eneste gyldige valuta er tomme flasker og bokser. Dette gjør at vi ikke får samme utvalg som andre studier ved at problemspillere blir selektert bort.

Pantelotteriet skiller seg også fra andre lotterier ved å tydelig merke at lotteriets overskudd går til Røde Kors. Som diskutert i Kapittel 2, kan gleden av å bidra til et samfunnsnyttig formål være en driver for konsum av veldedige lotterier. At pantelotteriets overskudd går til Røde Kors kan påvirke hvorvidt pantelotteriet betraktes på lik linje som andre lotterier, eller om kjøp av lodd blir sett på som en veldedig donasjon. Dersom dette er tilfellet, kan økonomiske teorier om altruistisk motivert spilling tilby et teoretisk rammeverk som forklarer den progressive skattefordelingen til pantelotteriet.

Både Suits-indeksene og inntektselastisiteten viser at pantelotteriet gjennom utvalgsperioden er progressivt beskattende, og at denne skattefordelingen bli mer proporsjonal gjennom utvalgsperioden. Gitt at pantelotteriet er et relativt nytt lotteri og at de estimerte fordelingseffektene endres gjennom hele utvalgsperioden, kan vi ikke anta at pantelotteriet vil være progressivt beskattende over tid.

Pantelotteriet bør med sin unike betalingsform ansees som et spesialtilfelle, og vi må være ytterst forsiktige med å generalisere resultatene denne oppgaven finner. Selv om de andre lotteriene drevet av samfunnsnyttige og humanitære organisasjoner må følge samme tekniske krav og retningslinjer som pantelotteriet, kan vi ikke anta at denne oppgavens resultater vil la seg reproducere for disse lotteriene. En mer eksplisitt sammenlikning av fordelingseffektene til pantelotteriet og andre lotterier drevet av samfunnsnyttige organisasjoner er nødvendig for å forstå om det er den unike betalingsformen som gir de spesielle fordelingseffektene av pantelotteriet.



## Referanser

- Anderson, J. E., Roy, A. G. & Shoemaker, P. A. (2003), 'Confidence intervals for the suits index', *National Tax Journal* pp. 81–90.
- Andreoni, J. (1990), 'Impure altruism and donations to public goods: A theory of warm-glow giving', *The economic journal* **100**(401), 464–477.
- Baker, R., Forrest, D. & Perez, L. (2016), 'Modelling regional lottery sales: methodological issues and a case study from Spain', *Papers in Regional Science* **95**(S1).
- Beckert, J. & Lutter, M. (2008), 'The inequality of fair play: lottery gambling and social stratification in Germany', *European sociological review* **25**(4), 475–488.
- Bernoulli, D. (1954), 'Exposition of a new theory on the measurement of risk', *Econometrica: Journal of the Econometric Society* pp. 23–36.
- Borge, L.-E. (2003), 'State lotteries in Norway, progressive or regressive taxation?', *Unpublished Manuscript*.
- Brinner, R. E. & Clotfelter, C. T. (1975), 'An economic appraisal of state lotteries', *National Tax Journal* pp. 395–404.
- Cameron, A. C. & Miller, D. L. (2015), 'A practitioner's guide to cluster-robust inference', *Journal of Human Resources* **50**(2), 317–372.
- Clotfelter, C. & Cook, P. J. (1987), 'Implicit taxation in lottery finance', *National Tax Journal* pp. 533–546.
- Clotfelter, C. T. (1979), 'On the regressivity of state-operated numbers games', *National Tax Journal* **32**(4), 543–548.
- Clotfelter, C. T. & Cook, P. J. (1990), 'On the economics of state lotteries', *Journal of Economic Perspectives* **4**(4), 105–119.
- Conlisk, J. (1993), 'The utility of gambling', *Journal of Risk and Uncertainty* **6**(3), 255–275.
- Doroghazi, R. M. (2009), Define specific goals, in 'The Physician's Guide to Investing', Springer, pp. 47–49.

- Forrest, D., Simmons, R. & Chesters, N. (2002), 'Buying a dream: Alternative models of demand for lotto', *Economic Inquiry* **40**(3), 485–496.
- Friedman, M. & Savage, L. J. (1948), 'The utility analysis of choices involving risk', *Journal of political Economy* **56**(4), 279–304.
- Frode Alfnes og Arne Dulsrud (2016), 'Vareutvalg av mat og drikke i norske dagligvarebutikker'.
- URL:** <https://fil.forbrukerradet.no/wp-content/uploads/2016/11/oppdragsrapport-nr-7-vareutvalg-av-mat-og-drikke-i-norske-dagligvarebutikker.pdf>
- Garrett, T. A. (2001), 'An international comparison and analysis of lotteries and the distribution of lottery expenditures', *International Review of Applied Economics* **15**(2), 213–227.
- Garrett, T. A. & Sobel, R. S. (2004), 'State lottery revenue: The importance of game characteristics', *Public Finance Review* **32**(3), 313–330.
- Gneezy, U. et al. (2013), *TheWhy Axis: Hidden Motives and the Undiscovered Economics of Everyday Life*, PublicAffairs, New York.
- Griffiths, M. D. & Wood, R. T. (1999), 'Lottery gambling and addiction: An overview of european research', *Report compiled for the Association of European National Lotteries (AELLE) Lausanne, Switzerland. Nottingham: Psychology Division, Nottingham Trent University. Mimeo.*
- Guryan, J. & Kearney, M. S. (2005), Lucky stores, gambling, and addiction: Empirical evidence from state lottery sales, Technical report, National Bureau of Economic Research.
- Hansen, A. (1995), 'The tax incidence of the colorado state lottery instant game', *Public Finance Quarterly* **23**(3), 385–398.
- Heavey, J. F. (1978), 'The incidence of state lottery taxes', *Public Finance Quarterly* **6**(4), 415–426.
- Hjorthol, R., Engebretsen, Ø. & Uteng, T. P. (2014), *Den nasjonale reisevaneundersøkelsen 2013/14: nøkkelrapport*, Transportøkonomisk institutt.
- Kahneman, D. & Tversky, A. (1979), 'Prospect theory: An analysis of decisions under risk', *Econometrica* **47**, 278.

Kocher, M. G., Krawczyk, M. & van Winden, F. (2014), ‘let me dream on!’ anticipatory emotions and preference for timing in lotteries’, *Journal of Economic Behavior & Organization* **98**, 29–40.

Langer, E. J. (1975), ‘The illusion of control.’, *Journal of personality and social psychology* **32**(2), 311.

Lavik og Jacobsen (2015), ‘Endringer i dagligvarmarkedet på 2000-tallet. handlevaner og holdninger’.

**URL:** [http://www.hioa.no/extension/hioa/design/hioa/images/sifo/files/file80415\\_oppdragsrapportnr.2-2015\\_handlevaner.pdf](http://www.hioa.no/extension/hioa/design/hioa/images/sifo/files/file80415_oppdragsrapportnr.2-2015_handlevaner.pdf)

Levy, H. (2015), *Stochastic dominance: Investment decision making under uncertainty*, Springer.

Lotteri og stiftelsestilsynet (2016), ‘Pengespelstatistikk kernel description’.

**URL:** <https://lottstift.no/pengespel/pengespelstatistikk/>

Markowitz, H. (1952), ‘The utility of wealth’, *Journal of political Economy* **60**(2), 151–158.

Mikesell, J. L. (1989), ‘A note on the changing incidence of state lottery finance’, *Social Science Quarterly* **70**(2), 513.

Mikesell, J. L. (1994), ‘State lottery sales and economic activity’, *National Tax Journal* pp. 165–171.

Miyazaki, A. D., Hansen, A. & Sprott, D. E. (1998), ‘A longitudinal analysis of income-based tax regressivity of state-sponsored lotteries’, *Journal of Public Policy & Marketing* pp. 161–172.

Norsk Tipping (2013), ‘Spillregler’.

**URL:** <https://www.norsk-tipping.no/spillregler/attachment/76949?ts=1438b2b3bd8>

Oster, E. (2004), ‘Are all lotteries regressive? evidence from the powerball’, *National Tax Journal* pp. 179–187.

Pallesen, S., Hanss, D., Mentzoni, R. A., Molde, H. & Morken, A. (2014), ‘Omfang av penge- og dataspillproblemer i norge 2013’, *Rapport for Lotteri og stiftelsestilsynet, Universitetet i Bergen, Institutt for samfunnspsykologi*.

- Perez, L. & Humphreys, B. (2013), 'The 'who and why' of lottery: empirical highlights from the seminal economic literature', *Journal of Economic Surveys* **27**(5), 915–940.
- Piketty, T. & Saez, E. (2003), 'Income inequality in the united states, 1913–1998', *The Quarterly journal of economics* **118**(1), 1–41.
- Price, D. I. & Novak, E. S. (2000), 'The income redistribution effects of texas state lottery games', *Public Finance Review* **28**(1), 82–92.
- Psacharopoulos, G. & Patrinos, H. A. (2004), 'Returns to investment in education: a further update', *Education economics* **12**(2), 111–134.
- Raaum, O., Aabø, T. E. & Karterud, T. (1999), 'Utdanning og livsinntekt i norge', *Oslo: Frisch-senteret*.
- Skirbekk, V. (2004), 'Age and individual productivity: A literature survey', *Vienna yearbook of population research* pp. 133–153.
- Spiro, M. H. (1974), 'On the tax incidence of the pennsylvania lottery', *National Tax Journal* pp. 57–61.
- Stanfield, B. & Kearney, M. S. (2008), 'Gambling at lucky stores: Empirical evidence from state lottery sales', *The American Economic Review* **98**(1), 458–473.
- Statistisk sentralbyrå (2018), 'Inntektsfordelingen belyst ved ulike ulikhetsmål. inntekt etter skatt, etter person, statistikkvariabel og år', <https://www.ssb.no/statbank/table/07756/tableViewLayout1/?rxid=c8ae23de-235b-4a40-82c6-0377055a4852>.
- Stranahan, H. & Borg, M. O. (1998), 'Horizontal equity implications of the lottery tax', *National Tax Journal* pp. 71–82.
- Suits, D. B. (1977), 'Measurement of tax progressivity', *The American Economic Review* **67**(4), 747–752.
- Titmuss, R. M. et al. (1970), 'The gift relationship. from human blood to social policy.', *The gift relationship. From human blood to social policy*.
- Tversky, A. & Kahneman, D. (1992), 'Advances in prospect theory: Cumulative representation of uncertainty', *Journal of Risk and uncertainty* **5**(4), 297–323.

Varian, H. R. (2010), *Intermediate Microeconomics; a modern approach*, WW Norton & Company.

Walker, I. (1998), 'The economic analysis of lotteries', *Economic policy* **13**(27), 358–401.

Wiepking, P. & Bekkers, R. (2012), 'Who gives? a literature review of predictors of charitable giving. part two: Gender, family composition and income', *Voluntary Sector Review* **3**(2), 217–245.

Wooldridge, J. (2013), 'Introductory econometry, a modern approach (vol. 5th)', *Mason: South-Western Publishing Company*.

## A Appendiks

### A.1 Bilde av spilleregler på panteautomat

**PANTELLOTTERIET**

---

**Generelle opplysninger:**  
 Lotteri til inntekt for Røde Kors med tillatelse gitt av Lotteri og stiftelsestilsynet for perioden 1. mars 2017 – 28. februar 2026. For å delta i Pantelotteriet brukes pant som innsats ved at det trykkes på Røde Kors-knappen 2 ganger for å bekrefte deltagelse i lotteriet.

Tillatt omsetningsgrense er på 120 millioner kroner per år, med tillatt antall solgte lodd på 240 millioner per år. Pris per lodd er 50 øre. Lotteriet har en gevinstandel lik 35%, 42 millioner kroner.

Alle gevinster inntil kr 1.000,- utbetales i kassen. Alle gevinster større enn kr 1.000,- utbetales gjennom lotteriselskapet, Norsk Pantelotteri AS, som kan kontaktes på telefon 815 52 222. Gevinster må avhentes innen 90 dager. Uavhentede gevinster tilfaller Røde Kors i sin helhet.

**Gevinstoversikt:**

12 premier à	1.000.000,-
6000 premier à	1.000,-
84000 premier à	100,-
312000 premier à	50,-


www.pantelotteriet.no

Figur 13: Bilde av informasjonsplakat klistert på panteautomaten ved Rema 1000 Bakklandet, Februar 2018.

## A.2 Korrelasjonsmatrise

Tabell 11: Korrelasjonsmatrise

Variable	Inntekt	Sysselsatte	18-49år	50-66år	67+år	IngenH.Utdanning	LavH.Utdanning	Bostedstype
Inntekt	1.00							
Sysselsatte	0.60	1.00						
18-49år	0.59	0.35	1.00					
50-66år	-0.64	-0.47	-0.90	1.00				
67+år	-0.69	-0.51	-0.83	0.84	1.00			
IngenH.Utdanning	-0.75	-0.31	-0.77	0.67	0.61	1.00		
LavH.Utdanning	0.72	0.29	0.70	-0.62	-0.58	-0.98	1.00	
Bostedstype	0.55	0.02	0.64	-0.59	-0.59	-0.67	0.69	1.00

## A.3 Variasjon mellom og innad i tverrsnittene

		Gjennomsnitt	Standardavvik
Spillandel	Overall	.084	.035
	Between		.034
	Within		.011
Inntekt	Overall	348 919	27 237
	Between		26 813
	Within		3 771
Sysselsetting i %	Overall	67.62	3.46
	Between		3.31
	Within		1.12
Andel 18-49år	Overall	0.43	.043
	Between		.043
	Within		.003
Andel uten høyere utdanning	Overall	0.23	.045
	Between		.045
	Within		.004
Andel bosatt i tettbygde strøk	Overall	0.82	.19
	Between		.19
	Within		.015
<i>N</i>		8330	

Tabell 12: Between- og Within-variasjon

## A.4 Komplette Tabeller

Tabell 13: Pooled OLS, Komplette tabell

	(1)	(2)	(3)	(4)	(5)	(6)
	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)
ln(Inntekt)	2.415*** (29.25)	3.633*** (31.02)	3.197*** (25.26)	2.106*** (11.07)	2.066*** (10.13)	2.016*** (10.17)
år2013	0.0817*** (15.47)	0.0580*** (10.71)	0.0540*** (9.23)	0.0705*** (11.29)	0.0698*** (10.87)	0.0769*** (12.16)
år2014	0.0851*** (10.31)	0.0577*** (6.85)	0.0462*** (4.54)	0.0707*** (6.67)	0.0707*** (6.50)	0.0987*** (9.09)
år2015	0.183*** (21.19)	0.0450*** (3.62)	0.0401** (2.67)	0.0798*** (5.09)	0.0817*** (4.88)	0.123*** (7.50)
ln(Syssellsetting)		-2.862*** (-15.74)	-2.443*** (-12.39)	-2.084*** (-10.70)	-2.031*** (-8.96)	-1.728*** (-7.82)
ln(18-49år)			1.843*** (9.65)	0.782*** (3.37)	0.774*** (3.31)	0.881*** (3.86)
ln(50-66år)			0.109 (0.61)	0.00538 (0.03)	0.0200 (0.11)	0.110 (0.63)
ln(66+år)			0.607*** (7.31)	0.264** (2.86)	0.264** (2.85)	0.256** (2.87)
ln(Ingen H.utdanning)				-0.780*** (-4.94)	-0.802*** (-4.62)	-0.762*** (-4.48)
ln(Kort H.utdanning)				-0.0206 (-0.24)	-0.0387 (-0.37)	-0.0860 (-0.84)
ln(Tettsted)					0.0179 (0.45)	-0.00350 (-0.09)
rimi						0.167*** (4.25)
kiwi						0.0674 (1.86)
rema						0.0672 (1.84)
ica						0.244*** (5.80)
coop						0.166*** (3.52)
ultra						0.179** (2.94)
spar						-0.0413 (-0.93)
centra						0.409*** (3.92)
meny						0.280*** (7.29)
Konstantledd -19.58***	-20.18*** (-31.82)					
		(-34.48)	(-23.21)	(-10.53)	(-10.19)	(-10.82)
N	8330	8317	8317	8317	8295	8295

*t*-verdier basert på clusterrobuste standardavvik i parentes.

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



Tabell 14: Random Effects, Komplett tabell

	(1)	(2)	(3)	(4)	(5)	(6)
	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)
ln(Inntekt)	2.105*** (5.65)	2.762*** (6.62)	2.436*** (8.12)	1.141*** (3.64)	1.058*** (3.33)	1.113*** (3.55)
år2013	0.0864*** (5.94)	0.0726*** (5.11)	0.0658*** (4.84)	0.0799*** (5.97)	0.0814*** (5.96)	0.0822*** (5.97)
år2014	0.115*** (5.74)	0.0949*** (5.00)	0.0717** (3.23)	0.0925*** (4.39)	0.0943*** (4.43)	0.0990*** (4.58)
år2015	0.219*** (9.47)	0.141*** (5.10)	0.115*** (3.99)	0.146*** (5.28)	0.153*** (5.34)	0.160*** (5.50)
ln(Syssellsetting)		-1.517*** (-4.27)	-1.270*** (-4.18)	-0.959*** (-3.32)	-0.819** (-2.67)	-0.754* (-2.55)
ln(18-49år)			1.672*** (4.00)	0.286 (0.63)	0.223 (0.47)	0.284 (0.61)
ln(50-66år)			-0.171 (-0.56)	-0.383 (-1.27)	-0.387 (-1.27)	-0.372 (-1.24)
ln(66+år)			0.680*** (3.79)	0.353 (1.87)	0.363 (1.89)	0.386* (2.03)
ln(Ingen H.utdanning)				-0.537 (-1.92)	-0.593* (-2.10)	-0.557* (-1.98)
ln(Kort H.utdanning)				0.428* (2.36)	0.373* (2.00)	0.343 (1.84)
ln(Tettsted)					0.0546 (1.08)	0.0299 (0.62)
Rimi						0.0668** (2.74)
Kiwi						-0.00733 (-0.18)
Rema						-0.0182 (-0.59)
Ica						0.0710** (2.61)
Coop						0.0559 (1.78)
Ultra						0.0449 (0.60)
Spar						-0.143*** (-3.43)
Centra						0.145*** (5.12)
Meny						0.130* (2.46)
Konstantledd	-29.54*** (-6.26)	-31.50*** (-7.13)	-25.89*** (-7.55)	-12.40*** (-3.56)	-12.05*** (-3.46)	-12.96*** (-3.76)
N	8330	8317	8317	8317	8295	8295

*t*-verdier basert på clusterrobuste standardavvik i parentes.

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

Tabell 15: Fixed Effects, Komplet tabell

	(1)	(2)	(3)	(4)	(5)	(6)
	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)
ln(Inntekt)	0.304 (0.52)	0.271 (0.39)	0.201 (0.31)	-0.922 (-1.66)	-0.882 (-1.56)	-0.894 (-1.56)
år2013	0.114*** (6.74)	0.114*** (6.00)	0.101*** (4.66)	0.0987*** (5.01)	0.0977*** (4.83)	0.0994*** (4.73)
år2014	0.160*** (6.66)	0.161*** (6.01)	0.121*** (3.41)	0.115*** (3.52)	0.112*** (3.36)	0.117** (3.24)
år2015	0.272*** (9.71)	0.276*** (6.01)	0.225*** (4.06)	0.208*** (4.29)	0.204*** (4.14)	0.212*** (3.94)
ln(Syssellsetting)		0.0824 (0.14)	0.273 (0.59)	0.309 (0.93)	0.339 (1.01)	0.364 (1.08)
ln(18-49år)			0.437 (0.44)	-0.853 (-1.26)	-0.766 (-1.11)	-0.736 (-1.08)
ln(50-66år)			-0.870 (-1.83)	-0.634 (-1.62)	-0.648 (-1.64)	-0.650 (-1.64)
ln(66+år)			0.772 (1.90)	0.544 (1.63)	0.570 (1.72)	0.573 (1.70)
ln(Ingen H.utdanning)				-0.351 (-0.60)	-0.411 (-0.70)	-0.193 (-0.28)
ln(Kort H.utdanning)				1.341*** (4.91)	1.398*** (5.00)	1.426*** (5.18)
ln(Tettsted)					-0.150 (-1.90)	-0.170* (-2.09)
rimi						0.0664** (2.67)
kiwi						0.0651 (1.87)
rema						0.194 (0.75)
ica						0.0412 (1.55)
coop						0.0601* (2.29)
ultra						-0.204* (-2.10)
spar						-0.126** (-3.04)
centra						-
meny						-0.109 (-1.36)
_cons	-6.584 (-0.88)	-6.507 (-0.85)	-5.856 (-0.88)	9.053 (1.36)	8.548 (1.26)	8.671 (1.27)
N	8330	8317	8317	8317	8295	8295

*t*-verdier basert på clusterrobuste standardavvik i parentes.

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

## A.5 Heterogenitet over tid

Tabell 16: Pooled OLS med Interaksjon

	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)
ln(Inntekt)	3.017*** (29.81)	4.221*** (31.17)	3.769*** (26.61)	2.691*** (13.70)	2.607*** (12.53)	2.569*** (12.69)
ln(Inntekt*2013)	-0.645*** (-9.92)	-0.622*** (-9.40)	-0.632*** (-9.80)	-0.628*** (-9.86)	-0.629*** (-9.85)	-0.612*** (-9.58)
ln(Inntekt*2014)	-0.840*** (-8.50)	-0.864*** (-8.75)	-0.838*** (-8.66)	-0.863*** (-8.98)	-0.879*** (-9.03)	-0.920*** (-9.50)
ln(Inntekt*2015)	-0.851*** (-7.94)	-0.802*** (-7.40)	-0.793*** (-7.35)	-0.934*** (-8.48)	-0.957*** (-8.53)	-1.021*** (-9.15)
Inkluderte variabler fra modellspesifikasjon:	(1)	(2)	(3)	(4)	(5)	(6)
Konstantledd	-41.14*** (-31.91)	-44.40*** (-32.93)	-37.47*** (-24.84)	-27.49*** (-13.81)	-27.01*** (-13.42)	-27.83*** (-14.15)
N	8330	8317	8317	8317	8295	8295
R <sup>2</sup>	0.24	0.30	0.35	0.36	0.36	0.40

t-verdier basert på clusterrobuste standardavvik i parentes.

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

## A.6 Robustehet MHT alternativ inntektsmål

Tabell 17: Gjennomsnittlig inntekt, Fixed Effects

	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)
ln(Gjennomsnittlig inntekt)	0.049 (0.16)	0.013 (0.04)	0.088 (0.25)	-0.149 (-0.48)	-0.108 (-0.34)	-0.116*** (-0.35)
Inkluderte variabler fra modellspesifikasjon:	(1)	(2)	(3)	(4)	(5)	(6)
Konstantledd	-3.34 (-0.84)	-3.48 (-0.89)	-4.64 (-1.20)	-0.21 (-0.06)	-0.67 (-0.20)	-0.65 (-0.19)
N	8330	8317	8317	8317	8295	8295
R <sup>2</sup>	0.45	0.45	0.46	0.47	0.48	0.48

t-verdier basert på clusterrobuste standardavvik i parentes.

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

Tabell 18: Gjennomsnittlig inntekt, Pooled OLS

	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)	ln(Spillandel)
ln(Gjennomsnittlig inntekt)	1.792*** (31.90)	2.244*** (32.14)	1.945*** (27.99)	1.577*** (12.45)	1.557*** (11.66)	1.497*** (11.58)
Inkluderte variabler fra modellspesifikasjon:	(1)	(2)	(3)	(4)	(5)	(6)
Konstantledd	-21.88*** (-35.64)	-25.77*** (-31.46)	-17.47*** (-20.43)	-13.76*** (-10.62)	-13.68*** (-10.51)	-14.23*** (-11.20)
N	8330	8317	8317	8317	8295	8295
R <sup>2</sup>	0.28	0.32	0.36	0.37	0.37	0.40

t-verdier basert på clusterrobuste standardavvik i parentes.

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