

HVEM BETALER FOR SVØMMEHALLENE?

EN ØKONOMETRISK ANALYSE AV UTGIFTER TIL PENGESPILL

WHO PAYS FOR PUBLIC SWIMMING POOLS?

LOTTERY EXPENDITURES IN NORWAY, AN ECONOMETRIC ANALYSIS

TERJE MOXNESS KORTNER

UNIVERSITETET FOR MILJØ- OG BIOVITENSKAP
INSTITUTT FOR ØKONOMI OG RESSURSFORVALTNING
MASTEROPPGAVE 30 STP. 2012



Forord

Denne oppgaven markerer avslutningen på to og et halvt år med studier ved Universitetet for miljø- og biovitenskap. Gjenstående svakheter ved oppgaven er mitt ansvar alene, men mange har bidratt positivt underveis i skriveprosessen.

En stor takk rettes til mine to veiledere, Olvar Bergland og Eirik Romstad, ved Institutt for økonomi og ressursforvaltning. Olvars korte og presise svar har vært til stor hjelp i en arbeidsprosess som tidvis har vært noe utflytende fra min side. Eirik har tatt seg tid til de lange diskusjonene, som tidsmessig har gått langt ut over hva jeg hadde forventet av en biveileder. Tusen takk til dere begge!

Takk til Erlend Dancke Sandorf for gode diskusjoner på kontoret og for utfyllende kommentarer til utkast. Takk også til dere andre medstudenter som har orket å høre på et endeløst mas om pengespill i lunsjpausene.

Takk til Hans Olav Fekjær for gode innspill tidlig i skriveprosessen.

Mastergraden ville ikke blitt fullført uten fleksibilitet fra min arbeidsgiver, Holmen / Kirkens Bymisjon. Takk til Vidar Nord og øvrige kollegaer for at dere har vist tålmodighet med en tidvis fraværende nattevakt. En spesiell takk rettes til Bjørn Louis Hjertnes, Nanny Weisser Eide og Marius Eide for å vise stor forståelse når jobbturnus har kollidert med studiearbeid.

Takk til kollega Merethe Lind for å ha vært et hyggelig reisefølge til Ås, og for gode innspill underveis i skrivearbeidet.

Til slutt retter jeg en stor takk til min venn, kjæreste og kone Stine Meløy, som har vært en uunnværlig støttespiller denne våren.

Ås, 4.5.2012

Terje Moxness Kortner

Sammendrag

Overskuddet fra statlige pengespill er med å finansiere fellesgoder i kultur-, organisasjons- og idrettssektoren i Norge, og en uttalt målsetning for Norsk Tipping er at de skal generere et størst mulig overskudd til slike formål. Tidligere forskning viser at utgifter til pengespill varierer mellom ulike sosioøkonomiske grupper, noe som resulterer i at konsumentene i ulik grad bidrar til overskuddsformålene. Flere studier finner at pengespill er mer utbredt i lavinntektsgrupper, og dersom vi anser netto utgifter til pengespill som en form for skattlegging kan pengespillene bryte med øvrig skattepolitikk.

Konsumentens inntekt eller velstandsnivå har stått som en sentral forklaringsvariabel i økonomisk teori om utgifter til pengespill. Tidligere empirisk forskning viser at andre sosioøkonomiske indikatorer som arbeidsstatus, utdanningsnivå, kjønn og alder er korrelert med pengespillutgifter. Oppgaven tar utgangspunkt i pengespill fra Norsk Tipping, som har om lag to tredjedeler av netto spilleomsetning for regulerte spill i Norge. Konkret forsøker oppgaven å besvare to spørsmål:

1. Hvilke sosioøkonomiske variabler kan forklare forskjeller i utgifter til pengespill hos norske konsumenter?
2. Flere fellesgoder finansieres av pengespill. Bryter dette med prinsipper om progressive eller proporsjonale skatter?

I de kvantitative analysene av spilleadferd benyttes to ulike datasett:

- (a) Kommuner. Datasettet inneholder total omsetning på spill i norske kommuner med utgangspunkt i reelle salgstall fra kommisjonærene, hentet fra Norsk Tippings offentlige statistikk. Medianinntekt for alle kommuner, samt en vektor av andre sosioøkonomiske variabler som kan korrelere med pengespill, er hentet ut av offentlig tilgjengelig kommunestatistikk fra Statistisk Sentralbyrå.
- (b) Husholdninger. Datasettet tar utgangspunkt i Forbruksundersøkelsen fra Statistisk Sentralbyrå, hvor et utvalg av norske husholdninger loggfører alle forbruksutgifter i en gitt periode. Utgifter til pengespill fra Norsk Tipping kodes i Forbruksundersøkelsen under en samlevariabel. Oppgaven benytter denne variabelen for spilleutgifter, og supplerer med en vektor av ulike sosioøkonomiske bakgrunnsvariabler for

husholdningen. I motsetning til kommunedata fra Norsk Tipping baseres spilleutgifter i Forbruksundersøkelsen på selvrapporing, noe som kan føre til seleksjonsproblematikk og underrapportering.

Spørsmål (1) besvares ved bruk av multippel regresjonsanalyse på begge datasett. Første regresjon benytter datasett (a), og estimerer spilleomsetning i kommunene på bakgrunn av medianinntekt og andre sosioøkonomiske forklaringsvariabler. Resultatene viser negativt parameterestimat for inntekt; dette kan tolkes som at innbyggerne i lavinntektskommuner spiller mer enn høyinntektskommuner, kontrollert for utdanningsnivå, arbeidsledighet, alders- og kjønnsammensetning med mer. Arbeidsledighet og høy andel med lavere utdanning i kommunen innvirker også positivt på spilleomsetning. Kulturutgifter og utgifter til sosialhjelp per innbygger er positivt korrelert med spilleomsetning. Sørlandskommuner har lavere spilleomsetning enn resten av landet.

Den andre regresjonen benytter husholdningsdata fra datasett (b), og estimerer utgifter til pengespill ved hjelp av Cragg-modellen. Dette er en hindermodell som skiller mellom det binære valget om å spille eller ikke spille, og mengdevalget gitt at husholdningen spiller. Modellen har lav predikasjonsverdi, og det ser ut til at utgifter til pengespill i liten grad er korrelert med inntekt i husholdningene.

Oppgaven diskuterer hvorvidt spørsmål (2) kan besvares med utgangspunkt i parameterestimatet for inntekt fra multippel regresjon. Oppgaven konkluderer med at dette ikke er korrekt metode til tross for at dette er fremgangsmåten i flere publiserte studier. Dette motiverer en regresjon med utgangspunkt i datasett (a), hvor medianinntekt settes som eneste uavhengige variabel. Sammenhengen mellom inntekt og spilleutgifter vil besvare spørsmål (2), også dersom denne sammenhengen skyldes bakenforliggende variabler, eksempelvis utdanning. En tovariats regresjon av medianinntekt på spilleomsetning gir et negativt parameterestimat på kommunenivå, som tolkes dit hen at innbyggerne i lavinntektskommuner i større grad bidrar til Norsk Tippings overskuddsformål enn innbyggerne i høyinntektskommuner. Visuelle plott bekrefter negativ sammenheng for hele inntektsintervallet, med unntak av én observasjon med høy innflytelse.

En stor andel av husholdningene i datasett (b) har ingen spilleutgifter, og det er store forskjeller i utgifter mellom husholdningene som deltar i pengespill. Utvalget deles inn i inntektskvintiler, og resultatene viser at husholdningene med lav inntekt bruker en større

andel av sine inntekter på pengespill. En beregnet Suits-indeks lik $-0,27$ styrker hypotesen om at pengespill er en form for regressiv skattlegging.

Summary

The profits from national lotteries are spent on common goods supplied by the cultural sector, sports sector and non-governmental organizations in Norway. Previous research shows that gaming expenses vary between socioeconomic groups, resulting in various contributions to these common goods from the consumers. Several studies find that lotteries and other forms of money gaming are more frequent in low-income groups. Assuming that national lotteries are a form of taxation, the resulting implicit tax from lotteries could be in conflict with a progressive tax system.

The current wealth or income level of the consumer is a key explanatory variable for gaming expenses in economic theory. Previous empirical research shows that socioeconomic variables such as employment status, education level, gender and age are correlated with gaming expenses. This thesis focuses on money games from Norsk Tipping, the state-run lottery company that accounts for two thirds of net turnover in the legal gaming market in Norway. The thesis addresses the following research questions:

1. Which socioeconomic variables can explain differences in gaming expenses of Norwegian consumers?
2. Several common goods are funded by the profits from national lotteries. Does this contradict progressive or proportional tax principles?

The thesis presents two data sets for quantitative analysis of gaming behavior:

- (a) Municipalities. The data set presents the observed gambling turnover in all municipalities, based on actual sales reported by Norsk Tipping. Statistics Norway provides publicly available statistics on the municipality level. Total gaming turnover is combined with median income and a vector of other explanatory variables for all municipalities.
- (b) Households. The data set consists of consumption data from a sample of Norwegian households, provided by Statistics Norway. Gaming expenses from said households in a given period are supplemented by a vector of background variables for every household. Consumption data from each household is self-reported and may suffer from underreporting and selection bias.

The thesis uses multiple regression analysis on both data sets in order to answer question (1). Starting with data set (a), median income and other explanatory variables are regressed on gaming turnover in the municipalities. The regression returns a negative parameter estimate for income, suggesting higher gaming levels among citizens in low-income municipalities when controlling for educational level, unemployment, age and gender effects. High unemployment and high proportions of low educated citizens give higher turnover estimates. High cultural and social welfare expenses in the municipality are positively correlated with gaming turnover, while lower gaming turnover is observed in southern Norway.

Secondly, using the Cragg model on data set (b), household income and other covariates are regressed on household gaming expenses. This hurdle model discriminates between the binary choice, i.e. whether to gamble or not, and the continuous one, i.e. the amount of money spent on a bet once the decision of betting has been taken. The model has low predictive power, and it looks like gaming expenses are poorly correlated with household income.

This thesis discusses whether the income parameter estimate from a multiple regression can answer question (2). The thesis concludes that this method is incorrect, despite being used in several published studies. This motivates a regression using data set (a) with median income as the only explanatory variable. The relationship between income and gaming expenses will answer question (2), also when facing a spurious correlation that is caused by other variables like education. A single regression of median income on gaming turnover gives a negative parameter estimate on the municipal level, suggesting that citizens in low-income municipalities contribute to the common goods to a higher degree than citizens in high-income municipalities. Excluding one influential observation, visual plots confirm a negative correlation for all income levels.

At the household level, a large fraction of the observations have zero gaming expenses, and there are substantial differences in gaming expenses for the households that participate in money games. Dividing the sample in income quintiles shows that households in the lower quintiles spend a larger portion of their income on money gaming. A calculated Suits-index of -0.27 strengthens the hypothesis of lotteries as a regressive form of taxation.

Innhold

Forord	1
Sammendrag	2
Summary	5
1 Innledning.....	9
1.1 Kort om bakgrunn og motivasjon.....	9
1.2 Problemstilling	10
1.3 Oppgavens oppbygging og hovedfunn.....	10
2 Bakgrunn	13
2.1 Pengespill i historien	13
2.2 Pengespill i Norge	14
3 Teori	17
3.1 Monetære verdier av pengespill	17
3.1.1 Nyttmaksimering	17
3.1.2 Usikkerhet.....	18
3.1.3 Friedman-Savage hypotesen.....	19
3.1.4 Kritikk av Friedman-Savage hypotesen	21
3.1.5 Prospektteori.....	24
3.2 Pengespill som underholdningsgode	26
3.3 Øvrige forklaringsvariabler	29
3.4 Pengespill som en form for skattlegging.....	31
3.4.1 Kvantifisering av skattebyrden.....	32
3.5 Motstridende formål	34
4 Data og metode.....	36
4.1 Metodevalg i tidligere forskning	39
4.2 Datamateriale.....	39
4.2.1 Omsetningstall, norske kommuner.....	40
4.2.2 Forbruksundersøkelsen.....	41
4.3 Økonometrisk modell	44
4.3.1 Kommuner.....	44

4.3.2 Husholdninger	46
5 Resultater og diskusjon	50
5.1 Resultater fra multippel regresjon, kommuner	50
5.2 Pengespill som implisitt skatt, kommunenivå	53
5.3 Resultater fra multippel regresjon, husholdninger	54
5.4 Pengespill som implisitt skatt, husholdningsnivå.....	57
5.5 Suits-indeksen.....	59
6 Avsluttende bemerkninger	61
6.1 Hovedfunn, kommuner.....	61
6.2 Hovedfunn, husholdninger	61
6.3 Sprik mellom kommune- og husholdningsdata	62
6.4 Videre forskning.....	63
7 Litteratur.....	64
Vedlegg 1	68
Vedlegg 2	69
Do-fil, husholdninger.	69
Do-fil, kommuner	71

1 Innledning

1.1 Kort om bakgrunn og motivasjon

I lotterier samles to eller flere aktører om et spill med usikkert utfall, de setter sin innsats, og lar tilfeldighetene avgjøre hvem som vinner. Slike sjansespill er sannsynligvis eldre enn sivilisasjonen, slik den oppstod i Mesopotamias byer for 7000 år siden. I dag har spillene utviklet seg til å bli varige samfunnsinstitusjoner med en betydelig økonomisk tyngde; globalt omsetter den lovlige spillindustrien for 335 milliarder amerikanske dollar per år (The Economist 2010). I Norge er Lottotrekningen et av landets mest populære TV-programmer¹ hvor hundretusener følger med for å se om de blir en av kveldens vinnere. Totalt bruker nordmenn nær 30 milliarder kroner på pengespill, og det statlige selskapet Norsk Tipping har om lag halvparten av brutto omsetning.²

Omfordeling er et sentralt begrep i analyser av pengespill (Beckert og Lutter 2009). For det første omfordeler pengespillene verdier mellom deltakerne, slik at noen ender opp som vinnere mens andre går tapende ut av spillet. For det andre omfordeler organiserte pengespill verdier mellom deltakere og arrangør, med den hensikt å skape et overskudd hos spillearrangøren. Denne kan være et privat selskap med ønske om profitt, eller et offentlig eiet selskap som benytter overskuddet i spillet til bestemte formål. Normalt kan vi anta at alle organiserte pengespill har en negativ forventningsverdi, altså at dette andre omfordelingsaspektet også er til stede. Dersom myndighetene har inntekter fra pengespill, enten fra overskuddet av et statlig spilltilbud eller gjennom skatteinntekter fra private pengespill som overgår det generelle skattenivået, kan pengespill også ha et tredje omfordelingsaspekt. Konsumenter med høye utgifter til pengespill vil da i større grad kunne bidra til statskassen enn konsumenter med lave pengespillutgifter, sammenliknet med øvrige skatteinntekter. Dette vil potensielt forskyve den totale skattebyrden mellom ulike grupper i samfunnet. Denne oppgaven vil konsentrere seg om å undersøke dette tredje aspektet av pengespill, det vil si hvordan spillene kan føre til omfordeling mellom ulike sosioøkonomiske grupper.

Mandatet til Skatteutvalget fra 2002 definerte de viktigste målene i den norske skatte- og avgiftspolitikken som "(...) å bidra til offentlige inntekter, utjevning av inntekter og levekår og til effektiv ressursutnyttelse" (NOU 2003: 11). Et progressivt skattesystem legger vekt på

¹ I uke 15 år 2012 var Lottotrekningen landets 9. mest sette enkeltprogram (TNS Gallup 2012).

² Se tabell 2.1.

utjevningensmålet, der progressive skatter er definert som skatter hvor skatteprosenten øker for høyere inntektsnivåer. Eksempelvis er inntektsskatten i Norge utformet som en progressiv skatt, grunnet minstefradrag på personinntekt og toppskatten. Forutsatt at andelen konsumutgifter av inntekt er lik for alle inntektsnivåer, vil en merverdiavgift eller omsetningsskatt være et eksempel på en proporsjonal skatt; alle inntektsnivåer vil ha lik skatteprosent.³ En vektet sum av de ulike skattene og avgiftene vil kunne svare på i hvor stor grad skattesystemet som en helhet er innrettet progressivt eller regressivt, eventuelt proporsjonalt. Samtidig er det også mulig å undersøke graden av progressivitet for hver enkelt skatt. Jeg ønsker å undersøke hvordan omsetningen på Norsk Tippings pengespill varierer for ulike inntektsnivåer i Norge.

1.2 Problemstilling

Konkret ønsker jeg å besvare følgende spørsmål:

1. Hvilke sosioøkonomiske variabler kan forklare forskjeller i utgifter til pengespill hos norske konsumenter?
2. Flere fellesgoder finansieres av pengespill. Bryter dette med prinsipper om progressive eller proporsjonale skatter?

Tidligere forskning viser at utgifter til pengespill varierer med ulike demografiske indikatorer som kjønn, alder, arbeidsstatus og inntekt (Beckert og Lutter 2009, Lund og Nordlund 2003, Pérez og Humphreys 2011). Ved hjelp av multippel regresjonsanalyse vil jeg undersøke de partielle korrelasjonene mellom slike forklaringsvariabler og utgifter til pengespill. Videre vil jeg diskutere hvorvidt parameterestimatet for inntekt i en slik multippel regresjon kan si noe om graden av progressivitet i lotteribeskatning, slik flere internasjonale studier antyder (Beckert og Lutter 2009, Pérez og Humphreys 2011, Worthington 2001).

1.3 Oppgavens oppbygging og hovedfunn

Umiddelbart kan det være vanskelig å forstå hvorfor ellers antatt nyttemaksimerende konsumenter betaler en positiv pris for en lottokupong med negativ forventningsverdi.

³ Den norske merverdiavgiften er utformet slik at en rekke varer og tjenester har lavere skattesats, eksempelvis matvarer. Hvorvidt matmomsen er et forsøk på omfordelingspolitikk eller landsbruksstøtte vil jeg ikke diskutere videre, men dersom vi antar at matvarer står for en høyere andel av utgiftene i husholdninger med lave inntekter, kan den norske merverdiavgiften ha et progressivt element. Samtidig kan forutsetningen om lik konsumandel i alle inntektsgrupper være problematisk, og empirisk forskning har vist at en flatskattesats på konsum kan være en regressiv skatt. (Suits 1977).

Samtidig observerer vi at pengespill er utbredt i de fleste kulturer. Jeg starter derfor oppgaven med et historisk tilbakeblikk på lotterier og andre pengespill i kapittel 2. Kapittel 3 ser på ulike økonomiske forklaringer på hvorfor en konsument kan velge å delta i spill som over tid påfører henne et økonomisk tap, og finner at konsumentens inntekt eller velstandsnivå står som en sentral forklaringsvariabel. Videre i kapitlet diskuterer jeg hvorvidt overskuddet i statlige pengespill kan anses som en form for skattlegging av konsumentene, og jeg forsøker å bestemme størrelsen på denne implisitte skatten.

Det har blitt gjort flere omfattende spørreundersøkelser om deltakelse i pengespill i Norge (Lund og Nordlund 2003, Pran og Ukkelberg 2010, Øren og Bakken 2007). Store utgifter til pengespill kan være forbundet med skyldfølelser for respondentene, som dermed vil kunne føre til underrapportering av spilleutgifter eller frafall av storspillere. Kapittel 4 presenterer og diskuterer to ulike datasett, et på kommune- og et på husholdningsnivå:

- (a) Norsk Tippings publiserer omsetningstall for alle norske kommuner, som kombinert med kommunedata fra Statistisk Sentralbyrå muliggjør en analyse uten frafallsproblematikk. Det er stor demografisk variasjon mellom norske kommuner, og jeg kan derfor gjøre en regresjonsanalyse med total kommuneomsetning som avhengig variabel.
- (b) Forbruksundersøkelsen fra Statistisk Sentralbyrå samler tverrsnittdata om konsumutgifter med utgangspunkt i selvrapportering fra norske husholdninger. Dette gir muligheten til å analysere på husholdningsnivå, men datasettet kan ha problemer med seleksjon og underrapportering.

I kapittel 5 estimerer jeg spilleomsetningen på bakgrunn av inntekts- og utdanningsnivå, arbeidsledighet i kommunen og andres sosioøkonomiske forklaringsvariabler, og finner negativ korrelasjon mellom kommuneinntekt og spilleomsetning. Dette kan skyldes utelatte variabler som samvarierer med inntektsnivået i kommunene, men det kan også indikere at pengespill er et mindreverdige gode i husholdningene.⁴ Kapittel 5 fortsetter med å estimere utgifter til pengespill i husholdningene. Her er bildet mindre tydelig, og jeg kan i liten grad forutbestemme pengespillutgifter på bakgrunn av valgte forklaringsvariabler. I husholdningene ser det ut til at pengespill i stor grad bestemmes av individuelle preferanser som ikke fanges opp av tilgjengelige data, og i liten grad varierer for ulike inntektsnivåer.

⁴ Et mindreverdige gode er i økonomisk forstand et gode hvor etterspørselen går ned for økt inntekt.

Spesielt bør leseren være oppmerksom på innledningen til kapittel 4 og avsnitt 4.1, hvor jeg diskuterer hvorvidt parameterestimatet for inntekt i en multippel regresjon på dobbel logaritmeform – det vil si inntektselastisiteten – kan si noe om graden av progressivitet i lotteribeskatning slik flere internasjonale studier antyder (Beckert og Lutter 2009, Pérez og Humphreys 2011, Worthington 2001). Det er gode argumenter for at inntektselastisiteter større enn 1 alene ikke bør tolkes som en progressiv lotteribeskatning, og jeg supplerer derfor med en tovariat regresjon i kapittel 5 for å besvare spørsmål (2). Videre beregner jeg Suits-indeksen (Suits 1977) for lotteribeskatning med utgangspunkt i husholdningsdata og finner $S = -0,27$, som indikerer at den implisitte skattebyrden gjennom statlige pengespill rammer regressivt i norske husholdninger. Kapittel 6 oppsummerer de viktigste funnene og diskuterer videre forskning i pengespillfeltet.

2 Bakgrunn

2.1 Pengespill i historien

Arkeologiske funn fra dagens nordlige Irak viser at terninger har vært brukt av mennesker i årtusener. Sekskantede terninger dateres til år 3000 før vår tidsregning, og før dette hadde talus-benet fra sauer blitt benyttet som en terning med fire utfall (Schwartz 2006: 8). Lotteriet som fordelingsmekanisme dukker også opp i Iliaden, hvor Poseidon forteller om da Zevs og hans brødre delte verden mellom seg:

*Alt blev deelt udi Tre, og Enhver sin Hæder annammed / Thi da vi kastede Lod, fik jeg
det graalige Havdyb / Evig til Boe, det taagede Mulm fik Hades til Eie / Zeus fik
Himlen den brede, i Luft og Skyer at herske*

Iliaden, 15. sang, 189 – 192.

Historikeren David Schwartz⁵ (2006) peker på at pengespill har vært utbredt blant urbefolkningsgrupper på alle kontinenter, og at de fleste religioner har tolerert pengespill i perioder. Under Hanukka leker mange jødiske barn fortsatt med en "dreidel", en snurrebass, hvor innsatsen er sjokolademynter eller ekte mynter. Klassisk indisk litteratur forteller om guden Shiva som spiller terning med sin kone Parvati, og vi finner ingen direkte forbud mot pengespill i bibelen. Islam er et unntak; pengespill omtales i Koranen som urent og "Satans verk" (Koranen, 5:90). De fleste større kinesiske byene hadde hus med organiserte dyrekamper 1000 år før vår tidsregning, og romerne var ivrige terningspillere; best illustrert av Julius Cæsars bemerkning *terningen er kastet* om hans krysning av Rubicon.

Pengespill kan være organisert direkte av deltakerne, eller det kan organiseres av en operatør, eksempelvis et statlig lotteriselskap eller private spillhus. Statlige lotterier dateres tilbake til keiser Augustus, som organiserte loddsalg til inntekt for vedlikehold av veier og bygninger i Roma. Schwartz (2006) antyder også at deler av den kinesiske mur ble finansiert av lotterier. Lotterier med pengepremier oppstod i byer som Ghent, Utrecht og Brugge allerede på 1400-tallet, hvor loddsalg ble organisert for å finansiere bymurene og andre forsvarsprosjekter. Nederlands Generaliteitsloterij, forløperen til nåtidens Staatsloterij, startet i 1726 og er et av

⁵ Schwartz er ansatt ved Center for Gaming Research ved University of Nevada, samt visepresident i spillskapet Santo Gaming, og er en kontroversiell figur i pengespillfeltet. Hans bok *"Roll the bones. The history of gambling."* (2006) som dette kapittelet henter stoff fra, gir en grundig oversikt over pengespillenes historie, men bør samtidig leses i lys av Schwartz' økonomiske forbindelser til spillindustrien.

de første eksemplene på statlige pengespill slik vi kjenner de i dag. Det første kjente offentlige spillhuset i den vestlige verden var Venezias Ridotto, som ble åpnet av byens myndigheter i 1638 for å regulere den illegale spillvirksomheten. Høye innsatser og en formell kleskode gjorde at det primært var byens bedrestilte som benyttet seg av Ridottoens spilltilbud, som bestod av både kortspill og loddtrekninger. I samme århundre la Pierre de Fermat og Blaise Pascal grunnlaget for moderne sannsynlighetsteori (Katz 1998: 448), og tilfeldighetene som tidligere rådet i spillhusene kunne nå settes i en vitenskapelig og forutsigbar ramme.⁶ Pascal konstruerte også ruletthjulet, som kort tid etter ble fast inventar i franske spillhus.

Parallelt med framveksten av offentlige spilltilbud eksisterte private alternativer. Lovligheten av disse varierte mellom land og i tid, men generelt forsøkte Europas myndigheter å begrense omfanget av pengespill. Samtidig så man både behovet for å regulere aktiviteten i ordnede former, slik Norsk Tipping i dag beskriver som "(...) å kanalisere folks spillelyst til et ansvarlig tilbud" (Norsk Tipping 2009: 2). I tillegg kunne pengespill være en enkel måte å finansiere offentlige prosjekter – i stedet for en påtvunget skattebyrde fikk befolkningen et spilltilbud som de selv kunne avgjøre om de ville benytte seg av. Eksempelvis er både Columbia University og British Museum bygget med delfinansiering fra lotterier.

2.2 Pengespill i Norge

Allerede i 1719 introduserte den dansk-norske kongen Frederik IV et lotteri til inntekt for foreldreløse barn, med store gårdsbruk som premier. Pengelotteriet var Norges første statlige pengespill, opprettet i 1912. Hestespill ble innført i 1927 etter påtrykk fra Bondepartiet, og humanitære organisasjoner som Røde Kors fikk tillatelse til å drive spilleautomater fra 1930-årene (Fekjær 2002: 30). I 1946 ble Norsk Tipping vedtatt opprettet, og første spilleomgang kom i 1948 med tolv engelske fotballkamper på tippekupongen. Selskapet hadde kun ukentlig fotballtipping frem til 1986, da Lotto ble lansert etter modell fra tilsvarende europeiske lotterier. Lottotrekningen fikk fast sendetid på NRK, og spillet bidro til en dobling av Norsk Tippings omsetning på fem år.

⁶ Girolamo Cardano hadde forsøkt dette et århundre tidligere, men hans bok om sannsynlighetsregning ble ikke utgitt før et århundre etter hans død. Den norsk-amerikanske matematikeren Øystein Ore går langt i å erklære Cardano som "the inventor of the theory of probability" (Ore 1953: 50). Cardano var for øvrig en ivrig deltaker i pengespill, og forteller i *Liber de Ludo Aleae* – "Boken om sjansespill" – at et spill endte med at han satte klærne som sikkerhet, og tapte. (Ore 1953: 217) Matematikkprofessor Audun Holme forteller at også Cardanos yngste sønn spilte bort alt han eide, "inklusive klærne sine" (Holme 2004: 216).

18-årgrensen for pengespill ble fjernet i 1985.⁷ Med Lotto og videre utover 1990-tallet endret spillmarkedet seg i Norge; Norsk Tipping lanserte nye spill og landet fikk en kraftig vekst i antall spilleautomater. De nye automatene godtok sedler opp til 500 kr, og fra 1990 til 1999 økte automatometningen med 4700 % (NOU 1997, Fekjær 2002). Ved årtusenskiftet oppsummerte overlege Hans Olav Fekjær at den norske automatpolitikken var Europas mest liberale. "Det omfattende pengespillet i Norge skaper glede i organisasjoner og driver tusenvis av familier til desperasjon og oppløsning" (Fekjær 2002: 135). I 2008 kom et forbud mot spilleautomater, som i dag er erstattet av Norsk Tippings spillterminaler. Disse kan minne om 90-tallets spilleautomater, men har funksjoner som skal forhindre spilleproblemer, eksempelvis øvre tapsgrense, tvungen spillepause hver time og alderskontroll.

De siste ti årene har det regulerte spillmarkedet i Norge gradvis fått økt konkurranse fra internettspill fra utenlandske kommersielle aktører. I 2009 spilte anslagsvis 8 % av Norges voksne befolkning på nettspill, og Lotteritilsynet anslo bruttoomsætning på utenlandske nettspill til 4,7 milliarder kroner (Lotteri- og stiftelsestilsynet 2010: 16). Norske finansinstitusjoner fikk forbud mot å overføre penger til utenlandske spilloperatører i 2010, samtidig som Norsk Tipping stadig utvider sitt tilbud på internett for å gi utenlandske operatører konkurranse.

Tabell 2.1: Omsetning pengespill i Norge 2010. Millioner kroner. (Kilde: Lotteri- og stiftelsestilsynet 2010)

	Brutto, ekskludert gevinstutbetalinger	Andel, brutto	Netto, inkludert gevinstutbetalinger	Andel, netto
Norsk Tipping	13 823	55 %	5125	59 %
Norsk Rikstoto	3667	13 %	1205	14 %
Tallspillet Extra	927	4 %	463	5%
Bingo	5600	22 %	1400	16 %
Lotteri, eksklusiv Flax	500	2%	400	5%
Spill på skip	469	2%	90	1 %
Sum, regulerte spill	24 986	100%	8683	100%
Nettspill, utlandet*	4700		902	

*Antar lik utbetalingsprosent som for spill på skip

⁷ Norsk Tipping gjeninnførte 18-års grense for alle spill 10. februar 2011 (Norsk Tipping 2010).

Spillmarkedet i Norge er presentert i tabell 2.1 I 2010 spilte nordmenn for nær 30 milliarder norske kroner, hvorav de regulerte spillene stod for anslagsvis 25 milliarder. Netto omsetning, det vil si at gevinstutbetalinger er fratrukket innsatsene, var for de regulerte spillene på 8,7 milliarder kroner. Vi ser at spill på utenlandske nettsider står for kun 9 % av total spilleomsetning. I tillegg til egne spillprodukter er Norsk Tipping operatør for Extra-lotteriet. Med utgangspunkt i netto omsetning ser vi da at Norsk Tipping har om lag to tredjedeler av det regulerte pengespillmarkedet i Norge.

3 Teori

3.1 Monetære verdier av pengespill

Ethvert pengespill har en forventningsverdi E , som kan uttrykkes som

$$(3.1) \quad E(w_1, \dots, w_n; p_1, \dots, p_n) = (w_1 p_1 + w_2 p_2 + \dots + w_n p_n) \quad , p_1 + \dots + p_n = 1$$

hvor w_1, \dots, w_n er alle mulige utfall av spillet med sannsynlighet p_1, \dots, p_n . Matematikeren Daniel Bernoulli diskuterte i det 18. århundre hvorvidt verdien av prospekter kunne variere for ulike inntektsnivåer. Generelt antok Bernoulli avtakende marginalnytte av penger: "(...) [T]he utility, however, is dependent on the particular circumstances of the person making the estimate. Thus there is no doubt that a gain of one thousand ducats is more significant to a pauper than to a rich man though both gain the same amount." (Bernoulli 1954: 24). Han hevder at et usikkert lodd med forventningsverdi lik 10 000 dukater kan gi den samme nytten som en lavere, men sikker, inntekt. Dette kunne forklare eksistensen av forsikring, men Bernoullis teori om avtakende marginalnytte impliserer at rettferdige lotterier⁸ generelt gir konsumenter lavere nytte.

3.1.1 Nyttmaksimering

Nyklassisk konsumentteori⁹ antar at nyttenivået til en konsument er en funksjon av mengde goder som konsumeres,

$$(3.2) \quad U = U(x_1, x_2, \dots, x_n)$$

Konsumet av hvert enkelt gode avhenger av prisen på alle goder, samt budsjettstrangen til konsumenten, og kan uttrykkes med (Marshall'ske) etterspørselsfunksjoner:

$$(3.3) \quad x_i = x_i^m(p_1, p_2, \dots, p_n, M)$$

hvor M er inntekten til konsumenten.¹⁰ Hvis funksjonene $x_i^m(\cdot)$ fra (3.3) substitueres inn i nyttefunksjonen $U(\cdot)$ fra (3.2) får jeg den indirekte nyttefunksjonen,

⁸ Et rettferdig lotteri defineres som et lotteri hvor forventningsverdien er lik innsatsen.

⁹ Se for eksempel Silberberg og Suen (2001: 252-276) eller Varian (1992: 94-108).

¹⁰ M kan defineres som "total budget per time period" (Silverberg og Suen 2001: 253), altså total mengde penger tilgjengelig for konsumenten per tidsperiode. Jeg oversetter dette til inntekt uten videre presisering.

$$(3.4) \quad U^*(p_1, p_2, \dots, p_n, M) \equiv U(x_1^m(p_1, p_2, \dots, p_n, M), \dots, x_n^m(p_1, p_2, \dots, p_n, M))$$

Jeg antar at $U^*(\cdot)$ er ikke-fallende i M . Dersom vi også antar sterk monotonitet, det vil si at konsumenten foretrekker å ha minst like mye av alle goder og strengt mer av minst et gode, har vi

$$\frac{\partial U^*}{\partial M} > 0$$

det vil si at nyttenivået øker for økt inntekt (Varian 1992: 104).

I moderne konsumentteori er $U(\cdot)$ og den tilhørende indirekte nyttefunksjonen $U^*(\cdot)$ ordinale, det vil si at de kun kan rangere mellom knipper av goder, men funksjonsverdien uttrykker ikke noe kardinalt mål på nytte. Dermed blir Alfred Marshalls antakelse om avtakende marginalnytte av penger (Marshall 2009: 111-112) moderert av eksempelvis Silberberg og Suen (2001: 265): "(...)[T]he concept of *diminishing marginal utility* is irrelevant in modern economics." Samtidig finner vi fortsatt en antakelse om avtagende marginalnytte i flere mikroøkonomibøker, se for eksempel Varian (1992: 104).¹¹

3.1.2 Usikkerhet

Dersom konsumenten står overfor usikkerhet, kan vi uttrykke nyttefunksjonen som

$$(3.5) \quad U(W_1, \dots, W_n; p_1, \dots, p_n) = p_1 u(W_1) + p_2 u(W_2) + \dots + p_n u(W_n)$$

eller i tilfellet med to goder

$$(3.6) \quad U(W_1, W_2; p_1, p_2) = p_1 u(W_1) + p_2 u(W_2)$$

hvor W_1, W_2 , er utfallene med sannsynlighetene p_1 og p_2 . Funksjonen $u(\cdot)$ kalles en Von Neumann-Morgenstern nyttefunksjon, etter deres arbeid i beslutningsteori. Funksjonen $u(\cdot)$ antas å være kardinal, og den tilhørende indirekte nyttefunksjonen $u^*(\cdot)$ uttrykker derfor konsepter som avtakende, konstant og tiltakende marginalnytte av inntekt.

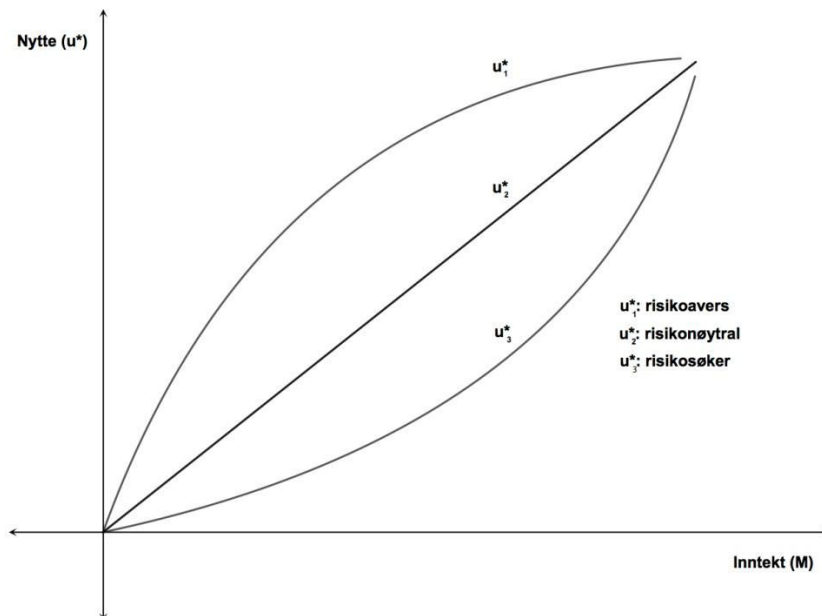
$$\frac{\partial^2 u^*(\cdot)}{\partial M^2} > 0 : \text{Konkav nyttefunksjon, avtakende nytte av inntekt}$$

¹¹ Figuren i Varian (1992:104) viser en nyttefunksjon av inntekt med avtakende marginalnytte.

$$\frac{\partial^2 u^*(\cdot)}{\partial M^2} = 0 : \text{Lineær nyttefunksjon, konstant nytte av inntekt}$$

$$\frac{\partial^2 u^*(\cdot)}{\partial M^2} < 0 : \text{Konveks nyttefunksjon, tiltakende nytte av inntekt}$$

En konsument med preferanse for risiko vil ha en konveks nyttefunksjon, en risikonøytral konsument vil ha en lineær nyttefunksjon, og en risikoavers konsument vil ha en konkav nyttefunksjon. Disse nyttefunksjonene er presentert i figur 3.1.



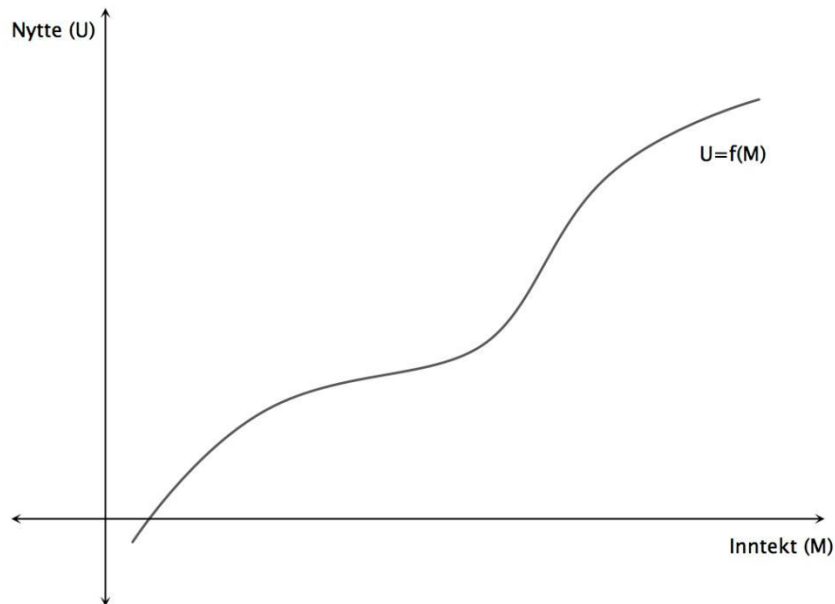
Figur 3.1: Risikopreferanser

Under antagelsen om avtakende marginalnytte av inntekt kan eksistensen av pengespill uten ferdighetselementer være vanskelig å forklare. Selv et rettferdig lotteri vil påføre spilleren et nyttetap, der et rettferdig lotteri defineres som et lotteri med forventningsverdi lik innsats.

3.1.3 Friedman-Savage hypotesen

Statlige pengespill som Lotto er helt uten ferdighetselementer, og har en negativ forventningsverdi siden innsatsene også skal dekke lotteriets driftskostnader samt overføringer til idrett, kultur og organisasjonsliv. Når konsumenter deltar i slike lotterier kan

dette tolkes som risikosøkende adferd. Samtidig kan vi også observere kjøp av forsikring i husholdninger som deltar i pengespill, hvor forsikringskjøp er et uttrykk for risikoavers adferd. Dermed kan ikke von Neumann-Morgenstein nyttefunksjoner fullt ut forklare eksistensen av pengespill.



Figur 3.2: Friedman-Savage hypotesen

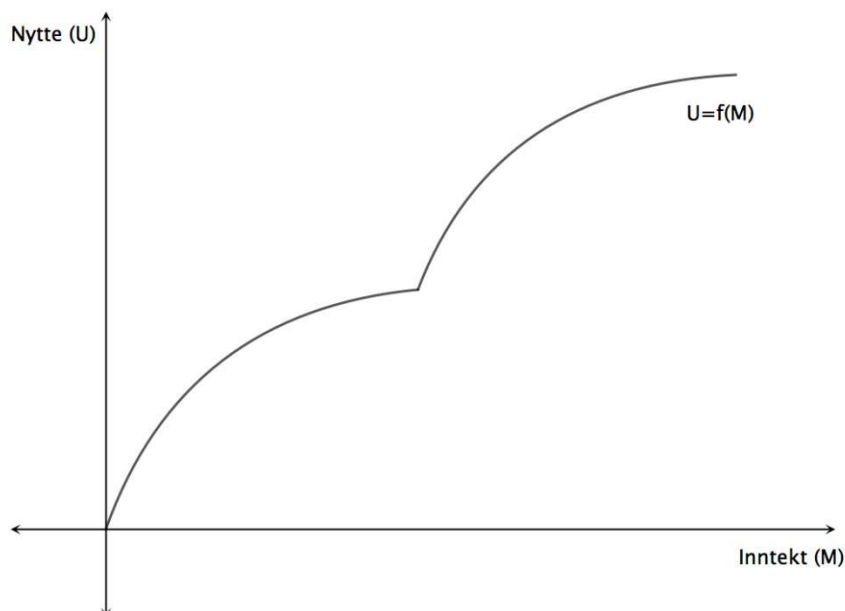
Friedman og Savage (1948) forlot antakelsen om generell avtakende marginalnytte av penger, og presenterte en nyttefunksjon som kombinerte konvekse og konkave segmenter, presentert i figur 3.2. Konkret mente de å kunne observere konsumenter med lave inntekter som både kjøpte forsikring og deltok i pengespill, og argumenterte for at en S-formet nyttefunksjon vil være konsistent med observert adferd. De tolket de konkave segmentene av nyttefunksjonen som to ulike sosioøkonomiske nivåer, og det konvekse segmentet som overgangen mellom de to nivåene. Konsumenter er risikoaverse for endringer innen sin nåværende klasse, men dersom en formuesøkning medfører et skifte inn i en ny samfunnsklasse, vil vi kunne observere økende marginalnytte:

An unskilled worker may prefer the certainty of an income about the same as that of the majority of unskilled workers to an actually fair gamble that at best would make him one of the most prosperous unskilled workers and at worst one of the least prosperous. Yet he may jump at an actuarially fair gamble that offers a small chance of lifting him out of the class of unskilled workers and into the "middle" or "upper" class, even though it is far more likely than the preceding gamble to make him one of the least prosperous unskilled workers.

Friedman og Savage 1948: 299

3.1.4 Kritikk av Friedman-Savage hypotesen

Ng Yew Kwang (1965) gir en alternativ forklaring på hvorfor vi kan observere pengespill blant konsumenter i lavinntektsgrupper. Han er inspirert av Friedman og Savage og deres nyttefunksjon, men erstatter det konvekse segmentet med et knekkpunkt, se figur 3.3.



Figur 3.3: Nyttefunksjoner med udelelige goder

Han argumenterer for at en del goder ikke kan deles opp: "In fact, consumption expenditures are not infinitely divisible; for example, while a person may choose between having one motor car or two, or even between having a Volkswagen and having a Mercedes 220, he cannot have half a car or one and one-third cars." (Kwang 1965: 30). Fattige familier deltar i lotterier og *number games*¹² fordi dette er den eneste reelle muligheten til å anskaffe en bil, et hus, en universitetsutdannelse eller starte et firma. Problemet med udelelighet minker for høyere inntektsnivåer, selv om vi også for relativt høy inntekt kan tenke oss udelelige konsumgoder som fritidsbolig ved sjøen og privatfly.

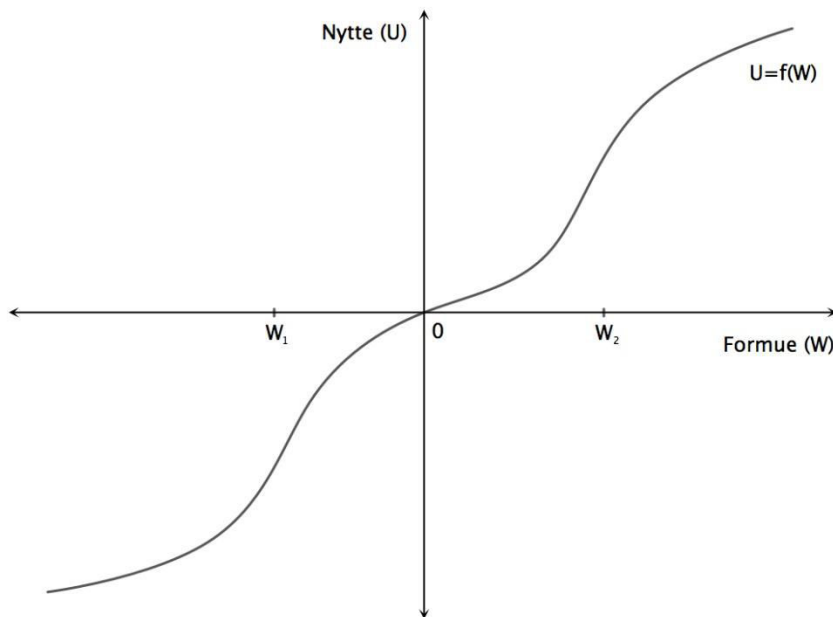
Harry Markowitz (1952) kritiserte Friedman-Savage hypotesen for å bryte med hva vi typisk observerer hos konsumenter. Dersom Friedman-Savage hypotesen skulle være riktig, medfører dette at konsumenter med middels inntekt bør være villige til å inngå store symmetriske veddemål (Markowitz 1952: 152). Dette mener Markowitz det er mangelfulle empiriske observasjoner for, og at slike veddemål heller tolkes som et tegn på irrasjonalitet. Samtidig viser han til at selv konsumenter med svært lave inntekter deltar i pengespill, samt at konsumenter med høye inntekter spiller rulett og er aktive i aksjemarkedet. Begge disse aktivitetene bryter med Friedman-Savage nyttefunksjonen i figur 3.2, da denne forutsetter risikoaversitet i disse gruppene.

Verken Friedman og Savage eller Markowitz har god empirisk bakgrunn for sine hypoteser, Markowitz viser kun til uformelle samtaler med bekjente. Basert på disse formulerer han en alternativ hypotese, hvor konsumentens nåværende formuesnivå danner et referansenivå. Han viser til at et flertall heller ønsker \$10 med 10 % sannsynlighet, enn \$1 med sikkerhet. Dette antallet synker for høyere veddemål med lik forventningsverdi: "Alle" velger \$1 000 000 med sikkerhet foran \$10 000 000 med 10 % sannsynlighet. Markowitz konkluderer med at nyttefunksjonen for formuesnivåer høyere enn nåværende formue først er konveks, deretter konkav.¹³ Han argumenterer videre for at konsumenter generelt har en preferanse for å tape \$1 med sikkerhet, framfor å risikere en gjeld på \$10 med 10 % sannsynlighet. For høyere gjeldsnivåer vil konsumenten foretrekke risiko – hun vil heller ta sjansen på å skylde \$10 000 000 med 10 % sannsynlighet framfor å skylde \$1 000 000 med sikkerhet. Dermed

¹² Se eksempelvis Spike Lee's "Malcolm X" (1992) for en presentasjon av *number games*, her med borgerrettighetsaktivisten Malcolm X som spilloperatør.

¹³ Merk at Friedman og Savage, samt Markowitz, konsekvent betegner et segment i en nyttefunksjon $U(\cdot)$ hvor $\partial^2 U / \partial x^2 < 0$ som *convex*, og segmentet hvor $\partial^2 U / \partial x^2 > 0$ som *concave*, se eksempelvis side 298, siste avsnitt i Friedman og Savage (1948). Denne misforståelsen kan ha bakgrunn i mengdelære (set theory), hvor en konkav funksjon danner en konveks mengde (convex set) *under* funksjonen. Likevel er det tydelig at forfatterne betegner selve nyttefunksjonen, og ikke mengden som denne avgrenser.

ender Markowitz opp med en nyttefunksjon presentert i figur 3.4, hvor nåværende formuesnivå er for $W = 0$. Vi har en funksjon som er konveks for $W < W_1$, verdier som kan tolkes som et stort formuestap fra nåværende formuesnivå. Funksjonen har vendepunkt for $W = W_1$, og er følgelig konkav i intervallet $(W_1, 0)$. Funksjonen er konveks i intervallet $(0, W_2)$, og konkav for $W > W_2$.

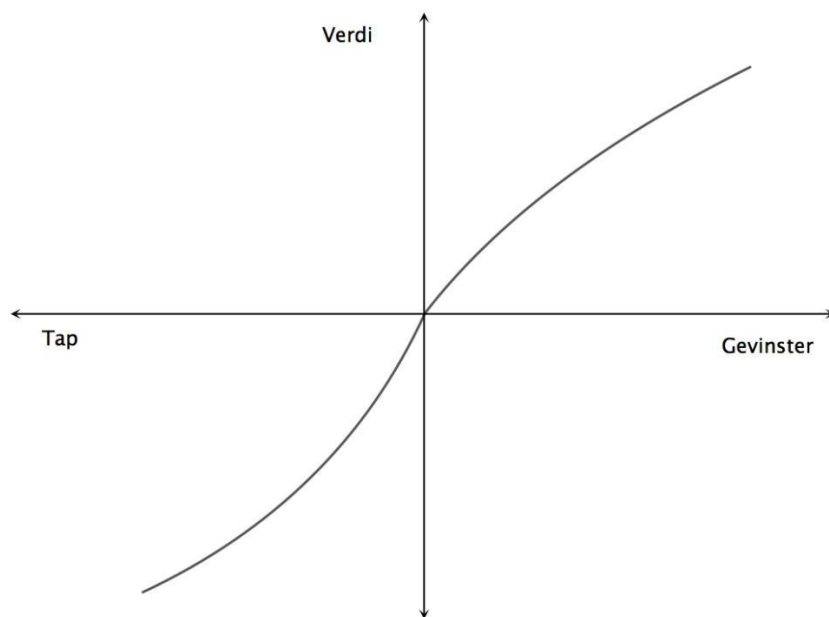


Figur 3.4: Markowitz' nyttefunksjon med referansenivå

Verdiene for vendepunktene W_1 og W_2 avhenger av personens nåværende inntektsnivå. Markowitz antar at en relativt rik konsument vil ha høyere (absolutt) verdi for disse punktene, eksempelvis vil hun også foretrekke \$1000 med 10 % sannsynlighet foran \$100 med sikkerhet. Avslutningsvis hevder han at konsumenter generelt unngår symmetriske veddemål, og at vi derfor har en brattere helning på kurven til venstre for origo: $|U(-x)| > U(x)$ for $x > 0$. Denne nyttefunksjonen kan forklare eksistensen av både lotterier og forsikring; grunnet konveksiteten til høyre for origo vil en konsument delta i pengespill, og konkaviteten til venstre for origo sikrer at samme konsument også er villig til å kjøpe forsikring. Hun vil ta store sjanser for et lite tap, mot en liten sjanse for en stor gevinst (Markowitz 1952: 155).

3.1.5 Prospektteori

Daniel Kahneman og Amos Tversky publiserte i 1979 en kritikk av teorier om maksimering av forventet nytte, og lanserte prospektteori som en forklaring på konsumentadferd under risiko. I likhet med Friedman og Savage, hevder de at von Neumann-Morgenstern nyttefunksjoner bryter med observert adferd, og viser til eksperimenter som bekrefter slike brudd. Konkret viser de at konsumenter generelt overvurderer lave sannsynligheter, eksempelvis sannsynligheten for husbrann eller for å vinne i et lotteri. Dette kaller de sikkerhetseffekten.¹⁴ Samtidig viser de at konsumentenes valg avhenger av om et alternativ presenteres som en gevinst eller et tap – eksempelvis foretrekker et flertall 3000 pund¹⁵ foran et lotteri med 80 % sannsynlighet for å vinne 4000 pund, og 20 % sannsynlighet ingen gevinst. Derimot foretrakk de samme respondentene et veddemål hvor man med 80 % sannsynlighet må betale 4000 pund, og 20 % sannsynlighet ikke må betale, foran en sikker gjeld på 3000 pund. Denne effekten kaller de refleksjonseffekten.¹⁶



Figur 3.5: Kahneman og Tverskys verdifunksjon

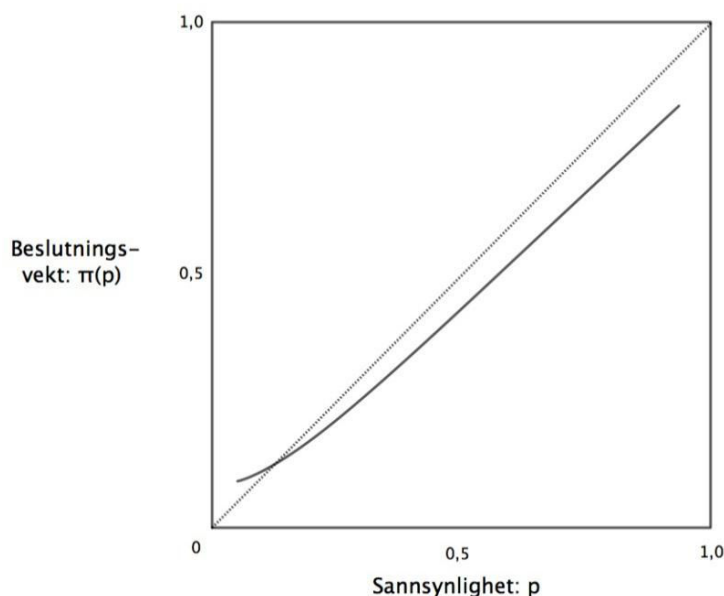
¹⁴ "Certainty effect". (Kahneman og Tversky 1979).

¹⁵ Israelske pund. 3000 I£ tilsvarte netto månedlig medianinntekt for en familie i Israel i 1979. (Kahneman og Tversky 1979: 264).

¹⁶ "Reflection effect".

Den observerte refleksjonseffekten gjør at Kahneman og Tversky ender opp med en nyttefunksjon presentert i figur 3.5 – kalt verdifunksjon for å skille den fra teorier om forventet nytte – som tar utgangspunkt i et referansenivå og avvik fra dette. Verdifunksjonen er generelt konkav for gevinster, og vanligvis konveks for tap, og har i motsetning til nyttefunksjonen til Markowitz (1952) ikke et initialt konvekst segment for gevinster og et initialt konkavt segment for tap. Vi ser da at forventet gevinst i et rettferdig lotteri vil være lavere enn forventet nytte av innsatsen, og følgelig vil ikke konsumenten delta i rettferdige eller urettferdige lotterier. Verdifunksjonen alene kan altså ikke forklare eksistensen av lotterier og forsikring, tvert i mot ser vi at konsumenten vil unngå både kjøp av forsikring og deltakelse i pengespill.

Samtidig vil sikkerhetseffekten virke i motsatt retning, da konsumenter generelt vil overvurdere lave sannsynligheter, eksempelvis risikoen for husbrann eller sannsynligheten for å få syv rette i Lotto. Prospektteorien multipliserer verdien av hvert utfall med en beslutningsvekt π^{17} , som er en funksjon av p , hvor p er sannsynlighetene for utfallene. $\pi(p)$ er voksende i p , med $\pi(0) = 0$ og $\pi(1) = 1$. Funksjonen er ikke kontinuerlig nær endepunktene 0 og 1.



Figur 3.6: Beslutningsvekt som funksjon av objektiv sannsynlighet

¹⁷ "Decision weight".

For utfall med svært lave sannsynligheter, som eksempelvis store gevinster i pengespill, vil altså verdien av lotteriet multipliseres med en beslutningsvekt som er høyere enn den faktiske sannsynligheten, og lotteriet kan være attraktivt også for konsumenter som generelt er risikoaverse for gevinster.

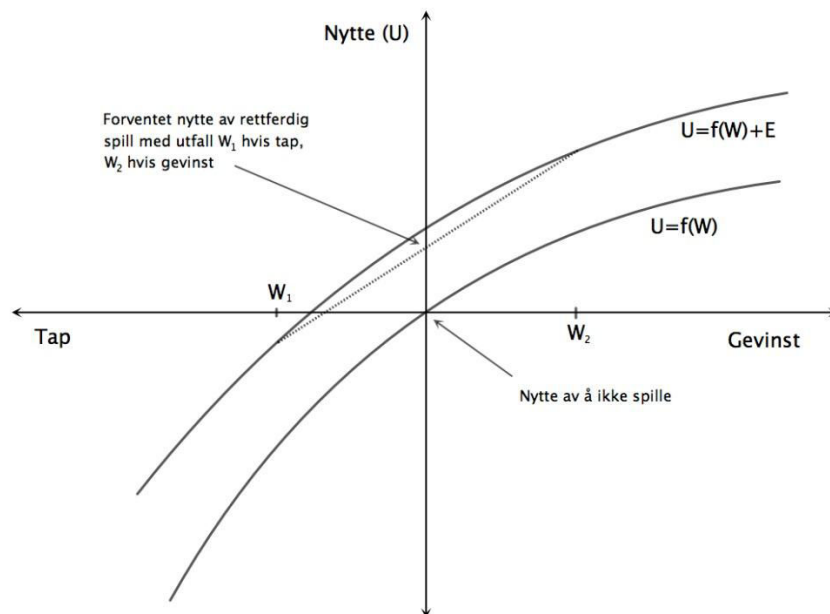
3.2 Pengespill som underholdningsgode

I tillegg til den rene monetære verdien av pengespill, er det også et underholdningsaspekt knyttet til spillene. Norsk Tipping beskriver seg selv som et selskap som "bidrar til å skape god underholdning og glede gjennom attraktive spill (...)" (Norsk Tipping 2009: 3). En SINTEF-rapport om pengespill av Lund og Nordlund (2003) bekrefter at underholdning står sentralt hos deltakere i norske pengespill. Respondentene bes om å oppgi de best likte egenskapene ved spillene de deltar i, og egenskapen "underholdende/spennende" er den best likte egenskapen både for problemspillere og spillere uten pengespillproblemer. Tilsvarende finner vi i Øren og Leistad (2010: 40): "Normalspillerne oppgir som oftest at årsaken til at de spiller er for spenningens skyld og for å vinne den store premien". Underholdningsaspektet står også sentralt for intervjuobjektene med pengespillproblemer; oppgitte årsaker til å spille i denne gruppen var blant annet "for spenningens skyld, for å vinne (lett-tjente) penger, en flukt fra hverdagen, tidsfordriv, for å få et 'kick' (...)" (Øren og Leistad 2010: 47).

Silberberg og Suen (2001: 409) konkluderer med at "[g]ambling can be explained by its entertainment value, consistent with the observation that people divide their stakes into small bets". De søker altså først og fremst til underholdningsaspektet for å forklare eksistensen av pengespill. Markowitz (1952: 157) diskuterer slike underholdningsforklaringer, og presenterer figur 3.7 hvor spillets underholdningsverdi E adderes til den monetære verdien. En konsument kan da ha en konkav nyttefunksjon, men likevel ha høyere nytte av å delta i pengespill.

Markowitz forkaster hypotesen om at underholdningsverdien kan forklare pengespill. Han viser til at konsumenten da vil maksimere nytte med svært lave innsatser, siden enhver økning i innsats vil gi konsumenten lavere forventet nytte. "In particular, when millionaires play poker together, they play for pennies; and no one will buy more than one lottery ticket. This contradicts observation." (Markowitz 1952: 157). Dersom Markowitz har rett i at innsatsene stiger når inntekten stiger, betyr ikke dette nødvendigvis at vi kan avvise at det finnes en underholdningsverdi knyttet til pengespill. Snarere kan denne observasjonen indikere at underholdningsverdien ikke bare handler om *selve spillet*, men at den også avhenger av

forholdet mellom potensielle gevinster G og spillerens inntekt M , hvor G er en lineær funksjon av innsatsen i spillet.¹⁸



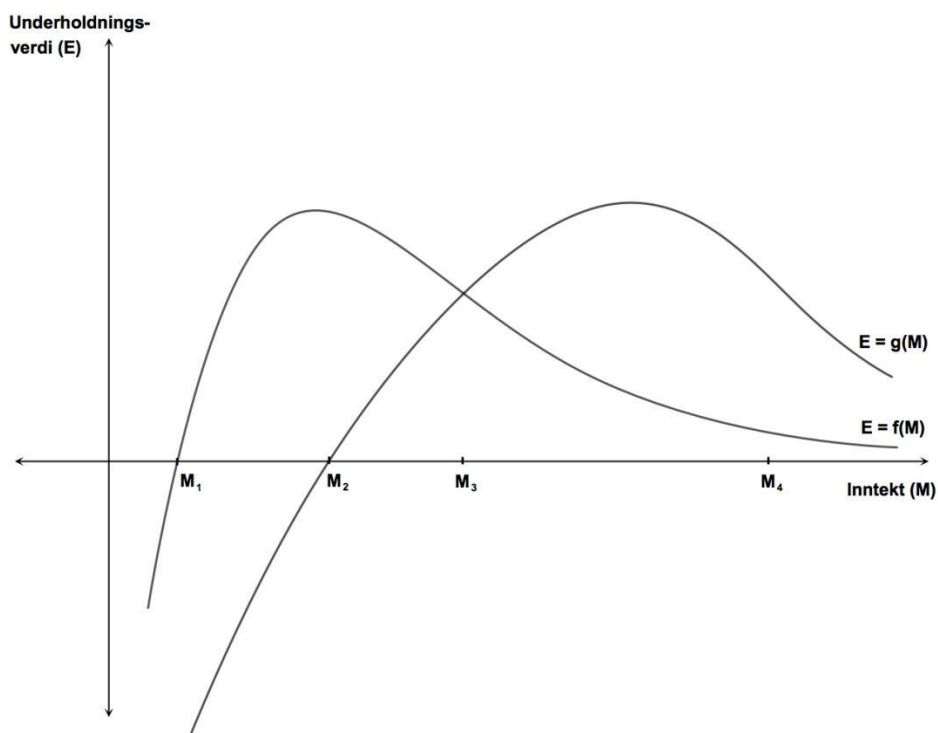
Figur 3.7: Underholdningsverdi av pengespill

Jeg antar at underholdningsverdien initialt stiger for høyere innsatser, fordi de høye innsatsene gir høyere potensielle gevinster. En millionær spiller heller poker med tusenlapper enn skillemynt, fordi hun relaterer den potensielle gevinsten G til sin inntekt M . Men på samme måte som en vanlig lønsmottaker normalt unngår å spille poker med tusenlapper, vil det også for Markowitz' millionærer eksistere et pokerbord hvor innsatsen er for høy sett i forhold til deres inntekt. Vi kan altså anta at underholdningsverdien av ethvert pengespill når et toppunkt for en gitt inntekt, for så å synke.

Figur 3.8 er en grafisk fremstilling av overstående resonnering, hvor de ulike funksjonene $f(M)$, $g(M)$, representerer lotterier med ulike innsats og tilhørende gevinster. Jeg antar at *selve*

¹⁸ Jeg antar her et rettferdig lotteri med lik innsats I for alle spillere, $G = n \cdot I$, hvor n er antall spillere.

spillet, uten innsatser, ikke har noen underholdningsverdi.¹⁹ $f(M)$ representerer underholdningsverdien til et lotteri med lave innsatser og tilhørende lave gevinster. Dette lotteriet vil ha negativ verdi for konsumenter med svært lave inntekter; eksempelvis kan et pokerbord hvor det spilles for hundrelapper ha negativ underholdningsverdi for barneskoleelever. For en konsument med ellers like preferanser, vil dette lotteriet framstå som mer attraktivt for stigende inntekt, og en lavtlønnet industriarbeider med inntekt $M > M_1$ har positiv underholdningsverdi ved dette pokerbordet. Derimot vil en velstående skipsreder med inntekt $M > M_4$ ha liten underholdning av dette spillet, siden det for henne i praksis er det samme som å spille for fyrstikker. Jeg antar derfor at $\lim_{M \rightarrow \infty} f(M) = 0$.



Figur 3.8: Hypotetisk underholdningsverdi for to lotterier med ulike innsatser

$g(M)$ representerer underholdningsverdien til et lotteri med høye innsatser og tilhørende høye gevinster. Dette lotteriet har negativ underholdningsverdi for alle konsumenter med $M < M_2$; det er altså mindre attraktivt for konsumenter med lave inntekter. Samtidig kan dette lotteriet være mer underholdende for høyinntektsgrupper; vi ser at for $M > M_3$ er lotteri g foretrukket

¹⁹ For poker vil underholdningsverdien av *selve spillet* for mange være positiv. Dette skifter kun kurvene i figur 3.8 vertikalt, og bryter ikke med det underliggende resonnementet.

foran lotteri f . Konsumenten med inntekt $M = M_4$ vil oppfatte lotteri g som mest underholdende, og hun vil kunne delta i spillet selv om den forventede monetære verdien er negativ. Funksjonene $f(M)$ og $g(M)$ sammenfaller med hva vi typisk kan observere; ulike inntektsgrupper velger ulike innsatser i ellers like spill.

3.3 Øvrige forklaringsvariabler

Jeg har nå diskutert ulike økonomiske forklaringer på eksistensen av pengespill, og ser at konsumentens inntektsnivå står som en sentral forklaringsvariabel på omfanget av spilldeltakelse. Prospektteorien (Kahneman og Tversky 1979), i likhet med Friedman-Savage hypotesen (Friedman og Savage 1948) og Markowitz' (1952) modifiseringer av denne, åpner for at konsumenters utgifter til pengespill har sammenheng med deres nåværende inntekt. Videre åpner Kwangs (1965) teori om udelelige goder for at konsumenter med lave inntekter spiller mer enn konsumenter med høyere inntekter. Teoretiske resonneringer om underholdningsverdien av pengespill indikerer at denne verdien varierer med inntekt; og at ulike lotterier kan appellere til ulike inntektsgrupper.

Pérez og Humphreys sammenfatter øvrige forklaringsvariabler på utgifter til pengespill som vi finner i internasjonale empiriske studier: "A common vector of explanatory variables, including age and age squared, income, gender, marital status, employment status, and the level of education, has been used in most previous empirical studies of consumer expenditure on gambling goods." (Pérez og Humphreys 2011: 560). Avslutningsvis vil jeg i dette kapitlet diskutere slike forklaringsvariabler med utgangspunkt i empiriske undersøkelser fra Norge.

Litteraturen jeg henviser til i dette underkapitlet baseres i første rekke på tovariate korrelasjoner, eksempelvis slik vi finner i Pran og Ukkelberg (2010: 5): "De demografiske mønstrene er vante, slik at menn, de over 50 år og de med lav utdanning oftere spiller lotto." Det menes her at pengespilldeltakelse korrelerer med kjønn, alder og utdanning; uten at forfatterne her forsøker å beskrive kausalitet eller korrigere for spuriøse sammenhenger. Med andre ord kan det godt tenkes at den typiske lottospilleren har lav utdanning *fordi* han er over 50 år, altså at variabelen alder både påvirker utdanningsnivå og deltakelse i pengespill. En multivariat tilnæringsmetode vil beskrives nærmere i kapittel 4, og foreløpig vil jeg benytte eksisterende litteratur om det norske pengespillmarkedet uten å problematisere hvorvidt variablene er reelle forklaringsvariabler eller om dette kun er spuriøse sammenhenger. Jeg vil

muligens finne at flere av forklaringsvariablene ikke har en selvstendig effekt på spilleomsetning, men åpner foreløpig for at alle observerte sammenhenger (delvis) kan skyldes variabelen i seg selv.

- **Kjønn:** Lund og Nordlund (2002: 58) finner at menn deltar i norske pengespill oftere enn kvinner, og viser til at det samme mønsteret er funnet for svenske pengespill. Dette bekreftes av Pran og Ukkelberg (2010: 39-50), som finner at menn spiller mer på alle norske pengespill unntatt Extra. Fekjær (2002: 75) hevder at spilleavhengighet er 2-3 ganger så vanlig hos menn som hos kvinner.
- **Alder:** Lund og Nordlund finner at andelen frekvente spillere²⁰ er stigende med alder. Samtidig ser det ut til at alvorlige spilleproblemer rammer hardere i yngre aldersgrupper; Pran og Ukkelberg (2010) finner at den typiske problemspilleren er en "ung mann", samt at internettspill ikke overraskende er mer vanlig blant den yngre del av befolkningen.
- **Utdanning:** Lund og Nordlund (2003: 57) finner at gruppen med universitets- og høyskoleutdanning har lavere andel frekvente spillere, og høyere andel ikke-spillere, sammenliknet med gruppene med lavere utdanning. Øren og Bakken (2007: 32) finner at det er lavere andel problem- og risikospillere blant respondentene med høyere utdanning. Pran og Ukkelberg (2010: 7, 39-45) finner at problemspillerne i større grad er lavt enn høyt utdannede, spesielt for spillene Lotto og Multix.
- **Yrkesdeltakelse:** Lund og Nordlund (2003: 57) finner små forskjeller i spildeltakelse mellom personer i arbeid og personer utenfor arbeidslivet. I den grad det er en forskjell, deltar personer i arbeid i større grad i pengespill. Pran og Ukkelberg (2010: 7) finner derimot større pengespillproblemer hos arbeidsløse og trygdede. Samtidig har vi grunn til å anta at tidkrevende pengespill som bingo og Multix i større grad tiltrekker personer utenfor arbeidslivet.
- **Bosted:** Pran og Ukkelberg (2010: 7) finner en overrepresentasjon av problemspillere i byene og i Nord-Norge. Øren og Bakken (2007: 32) finner derimot høyest andel problem- og risikospillere i tettsteder (ikke byer), og like store pengespillproblemer i Nord-Norge som på Sørlandet. Lund og Nordlund (2003: 57) finner at byene har størst andel ikke-spillere, samt at Nord-Norge har størst andel frekvente spillere.

²⁰ Denne gruppen har siste år spilt daglig på minst en spillgruppe, eller en til flere ganger i uka på minst to spillgrupper. Eksempler på spillgrupper er "lotteri", "hest" og "internett". Se Lund og Nordlund (2010: 53) for en uttømmende liste.

3.4 Pengespill som en form for skattlegging

Skatter kan defineres som tvungne overføringer til det offentlige som det ikke er noen direkte motytelser til (Grønn 2005: 232). Et eksempel på dette er merverdiavgift; norske konsumenter betaler 25 % avgift på vare- og tjenestekjøp til staten.²¹ Selvsagt kan konsumenten avstå fra forbruk, men med "tvungen" menes her at dersom hun velger går til anskaffelse av en vare, tvinges hun også til å betale avgiften. Samtidig får hun ikke noen direkte motytelse tilbake fra staten, skatten fordeles i sin helhet av de folkevalgte over statsbudsjettet.

Statlige pengespill er fritatt fra merverdiavgift. (Merverdiavgiftsloven 2009, § 3-14.) Samtidig går en betydelig andel av omsetningen til staten, som fordeler overskuddet mellom ulike overskuddsformål.

Tabell 3.1: Inntekter og utgifter 2009, Norsk Tipping. Millioner kroner. (Kilde: Norsk Tipping 2009)

		Andel
Spillinntekter	12575	100 %
Gevinster	7322	58 %
Overskuddsmottakere*	3235	26 %
Kostnader, drift**	2018	16 %

*Inkludert grasrotmottakere, ekskludert overføringer fra investeringsfond.

**Herunder lønn, pensjoner, skattetrekk, arbeidsgiveravgift, provisjon til kommisjonærer, avskrivninger; samt andre driftskostnader som drift av anlegg og maskiner, markedsføring, teleoperatørkostnader med mer. For utfyllende liste, se note 4 i Norsk Tippings regnskap (Norsk Tipping 2009: 66).

Som beskrevet i kapittel 2, har finansiering av samfunnsnyttige formål vært et av flere argumenter for å opprette og videreføre statlige pengespill. I dag beskriver Norsk Tipping denne delen av sitt samfunnsoppdrag som å "levere et størst mulig overskudd til overskuddsformålene" (Norsk Tipping 2009: 2). Overskuddsformålene til Norsk Tipping er i dag idrett (45,5 %), kultur (36,5 %) og samfunnsnyttige og humanitære organisasjoner

²¹ Dette er en forenkling; eksempelvis har matvarer lavere avgiftssats, og en rekke tjenester har fritak fra merverdiavgiften.

(18 %).²² Tabell 3.1 viser at 26 % av Norsk Tippings spilleinntekter i 2009 fordeles av myndighetene til nevnte formål.

Mange hevder at pengespill er en form for "frivillig skattlegging" (Fekjær 2002: 130), fordi konsumentene selv velger hvorvidt de deltar i pengespill. Et slikt begrep tilslører mer enn det oppklarer; konsumentene velger også hvorvidt de går til innkjøp av øvrige varer og tjenester, uten at noen vil kalle merverdiavgiften for en frivillig skatt. Enhver konsument med inntekt større enn innslagspunktet for toppskatten kan velge å si opp nåværende jobb og ta seg et lavere betalt arbeid, men ingen vil kalle toppskatten en frivillig skatt. Clotfelter og Cook (1987) benytter begrepet *implisitt skatt*²³ om overskuddsoverføringene fra (del-)statslotterier, og avviser at denne skattebyrden skiller seg vesentlig fra øvrige skatter. Sosiologene Jens Beckert og Mark Lutter (2009) argumenterer for at statlige pengespill er en form for skattlegging, og liknende argumenter finner jeg hos økonomene Levi Perez og Brad R. Humphreys: "Introducing lottery games gave government access to a new and substantial source of tax revenue." (Perez og Humphreys 2011: 552). Med utgangspunkt i definisjonen av skatt som en tvungen pengeoverføring til det offentlige uten direkte motytelser, er det gode argumenter for å se på overskuddet i statlige pengespill som en form for skattlegging.

3.4.1 Kvantifisering av skattebyrden

Kvantifiseringen av skattebyrden er mer utfordrende. I land hvor private spilloperatører skattlegges direkte av myndighetene, kan man ta utgangspunkt i slike pengeoverføringer. (Worthington 2001). Dette kompliseres i et land som Norge, hvor pengespill delvis er under direkte statlig eierskap (Norsk Tipping), delvis organiseres av en stiftelse (Norsk Rikstoto), og delvis opereres av kommersielle aktører (bingodrift). Den videre diskusjon av skattebyrden tar kun utgangspunkt i Norsk Tipping sine spilltjenester, altså et pengespill hvor staten både er eier og operatør, og samtidig fordeler overskuddet fra pengespillene.

Beckert og Lutter (2009: 476) definerer skatteraten i det tyske statslotteriet til å være overskuddet fra lotteriet som går tilbake til myndighetene; som i Tyskland er 39 % av omsetningen. Dette sammenlikner de med den tyske merverdiavgiften på 7 – 19 %, og konkluderer med at pengespill er et høyt²⁴ beskattet gode. En slik definisjon av skatteraten er

²² I tillegg kommer den såkalte Grasrotsandelen, hvor spillerne selv bestemmer seg for hvilket lag/forening som skal få støtte. I 2009 utgjorde dette totalt 211 mill kr. (Norsk Tipping 2009: 16).

²³ "Implicit tax". (Clotfelter og Cook 1987: 533)

²⁴ "Very highly". (Beckert og Lutter 2009: 475).

problematisk av flere grunner. For det første vil gevinstandelen variere sterkt mellom typen lotteri som blir tilbudt. Eksempelvis går 91 % av bruttoomsetningen til Norsk Tippings Multix-automater tilbake til spillerne i form av gevinster, mens gevinstandelen for norsk Lotto er 50 % (Norsk Tipping 2009: 4). Dermed vil overskuddets andel av bruttoomsetning være betydelig lavere for spilleautomater enn for lottospillene, uten at vi fra dette kan dra slutningen at Lotto er "høyere" beskattet. Videre kan statlige pengespill også ha ulike driftskostnader. Dersom driftskostnadene er svært høye, blir det mindre igjen som skal fordeles til gode formål. Alt annet likt, vil altså Beckert og Lutter med sin definisjon måtte anse et ineffektivt statslotteri som en lavere skattebyrde sammenliknet med et statslotteri som har lave driftskostnader.

For det tredje kan det være problematisk å benytte forventningsverdien til å si noe om skatteraten i spill som Lotto. Forventningsverdien er en middelvei – altså at man kan forvente å få tilbake 50 %²⁵ av innsatsen over tid:

$$(3.7) \quad E(\text{innsats} = I) = I * 0,5$$

Samtidig vet vi at fordelingen både har høy varians og at den er usymmetrisk; spill som Lotto er basert på at *mange* taper litt, for at *noen* skal vinne mye. Forventningsverdien sier noe om forventet gevinstutbetaling *over tid*, men utbetalingene vil ikke rekke å gå mot forventningsverdien for den enkelte spiller som har et begrenset antall leveår hun kan spille i. Det kan vises²⁶ at 9 av 10 spillere vil spille om de mindre gevinstene, og at denne gruppen, som verken vil vinne toppgevinsten (7 rette) eller andre gevinst (6 rette + 1 tilleggstill) har forventningsverdi lik

$$(3.8) \quad E(\text{innsats} = I \mid \text{ikke 7 rette, ikke 6 + 1 rette}) = I * 0,216$$

Nettopp fordi denne asymmetrien er en nødvendig betingelse for konsumenters deltakelse i pengespill som Lotto, kan toppgevinstene sees på som en form for driftskostnader. Tydeligere blir dette hvis gevinstene i et lotteri ikke er penger, men en gjenstand. Anta et lotteri organisert av et idrettslag, hvor den eneste gevinsten er en fin bil som kjøpes inn av idrettslaget²⁷. Lotteriets regnskap vil da føre loddsalget som inntekter, mens utgiftene vil

²⁵ Premieandelen i Lotto er 50 % av innsatsene. (Norsk Tipping 2009)

²⁶ Se vedlegg 1.

²⁷ Jeg forutsetter at idrettslaget må betale hele eller deler av gevinstens utsalgspris til bilforhandleren.

være driftskostnadene ved lotteriet, herunder innkjøp av gevinsten. Det vil være meningsløst å hevde at forventningsverdien i lotteriet er bilens verdi delt på antall lodd, selv om dette formelt sett er riktig. Hver deltaker i lotteriet har en utgift pålydende loddets innkjøpspris, som kan anses som støtten til idrettslaget. Åpenbart vil idrettslaget sitte igjen med et lavere beløp, siden alle former for pengeinnsamling har driftskostnader. Dersom laget velger å samle inn penger ved hjelp av et lotteri, vil de altså få økte kostnader sammenliknet med en ren pengeinnsamling, siden de må kjøpe inn gevinst(er). Samtidig vil også inntektene kunne stige, delvis grunnet sikkerhetseffekten diskutert i kapittel 3.1.5, og delvis grunnet lotteriets underholdningsverdi.

3.5 Motstridende formål

Statlige pengespill kan ha ulike formål, som er delvis motstridende. Eksempelvis beskriver Norsk Tipping sitt samfunnsoppdrag som tredelt; de skal (1) gi et ansvarlig spilltilbud, (2) et attraktivt spilltilbud og (3) generere et overskudd til gode formål. (Norsk Tipping 2009: 2). Norsk Tippings spillemonopol har altså et reguleringsaspekt. I likhet med rusmidler, er pengespill noe som samfunnet ser på som potensielt skadelig, og som dermed skal underlegges statlig kontroll. Isolert sett vil formål (1) føre til et statlig ønske om å begrense pengespill i befolkningen. Formål (2) handler om underholdningsaspektet til pengespill. Eksistensen av pengespill vitner om en etterspørsel etter dette godet blant konsumenter, og dette gjør at staten bør legge til rette for denne type tjenestetilbud på lik linje med andre goder som gir nytte i befolkningen. Dette formålet vil isolert sett medføre at pengespill tilbys i den grad det finnes etterspørsel etter godet. Formål (3) handler om å finansiere andre goder som befolkningen også etterspør, eksempelvis offentlige svømmehaller. Dette formålet vil isolert sett føre til et ønske om å utvide pengespilltilbudet, for å tilby befolkningen flere og bedre svømmehaller. Det er dette formålet som gir grunnlag for å anse pengespill som en form for skatteinnkreving.

Dermed er skatteinnkreving bare et av flere aspekter ved statlige pengespill, og dette gjør det vanskelig å fastsette "skatteraten" slik Beckert og Lutter (2009) forsøker å gjøre. Delvis grunnet argumentene presentert i avsnitt 3.4.1, og delvis fordi pengespillene også har andre formål. Når jeg i neste kapittel går over til den økonometriske analyse av utgifter til pengespill, minner jeg om at skatteinnkrevingen kun er et av flere samfunnsoppdrag for Norsk Tipping. Fotballtippingen som startet i 1948 kan til en viss grad anses som et forsøk på å regulere illegalt pengespill, mens bingospill på et eldrecenter i første rekke kan forklares ut i

fra beboernes behov for et underholdningstilbud. Flere hevder at veksten i norske pengespill fra midten av 1980-tallet bør sees i sammenheng med formål (3), altså et ønske om å få et høyere overskudd som kan fordeles til gode formål (Fekjær 2002).

4 Data og metode²⁸

Dette kapitlet vil gjøre rede for valg av metode og spesifisere en modell. Jeg vil videre presentere to datasett og diskutere mulige svakheter ved begge. Resultatene vil bli diskutert i neste kapittel.

Formålet med oppgaven er å besvare spørsmålene som ble stilt innledningsvis:

1. Hvilke sosioøkonomiske variabler kan forklare forskjeller i utgifter til pengespill hos norske konsumenter?
2. Flere fellesgoder finansieres av pengespill. Bryter dette med prinsipper om progressive eller proporsjonale skatter?

Spørsmål (1) vil kunne besvares av en lineær regresjonsanalyse, forutsatt at modellen i populasjonen er lineær i alle parametere, at populasjonsfunksjonen estimeres med et tilfeldig utvalg fra denne populasjonen og at feilleddet i den estimerte modellen har forventningsverdi lik null. Vi vil her forklare den avhengige variabelen y : *utgifter til statlige pengespill* med uavhengige variabler som x_1 : *disponibel inntekt*, x_2 : *utdannelse*, x_3 : *yrkesstatus* og liknende. I kapittel 3 har jeg diskutert slike forklaringsvariabler på utgifter til pengespill. Vi kan derfor med bakgrunn i økonomisk teori, samt empiriske funn om utgifter til pengespill, formulere en regresjonsfunksjon for populasjonen:

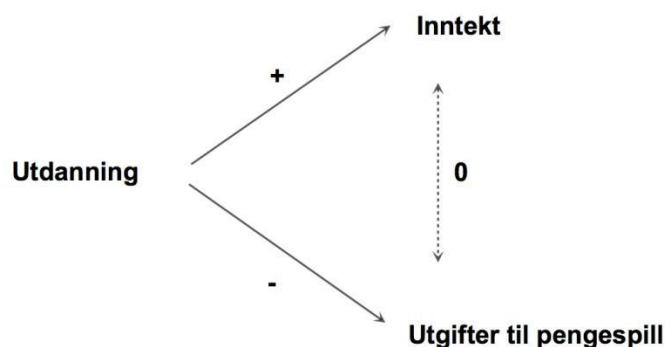
$$(4.1) \quad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + u$$

Hvor u er en uobserverbar tilfeldig variabel som kan ta positive eller negative verdier. Denne regresjonsfunksjonen kan være på lineær form som gitt i eksempelet over, eller uttrykkes med en annen funksjonell form. Parameterestimatene β_i vil angi effekten av hver enkelt variabel, kontrollert for øvrige variabler som er med i funksjonen. Dersom alle relevante variabler er tatt med, vil β_i måle den partielle effekten av forklaringsvariabelen på den avhengige variabelen, *ceteris paribus* – når alle øvrige variabler holdes konstant.

²⁸ En del av de data som er benyttet her er hentet fra "Forbruksundersøkelsen, 2006", samt "Forbruksundersøkelsen, 2005". Data er innsamlet av Statistisk sentralbyrå. Data er tilrettelagt og stilt til disposisjon i anonymisert form av Norsk samfunnsvitenskapelig datatjeneste AS (NSD). Verken Statistisk sentralbyrå eller NSD er ansvarlig for analysen av dataene eller de tolkninger som er gjort her.

Begrepene progressive og regressive skatter relateres normalt til totale skatteinnbetalinger for ulike inntektskvintiler i et land, se eksempelvis Suits (1977). Dersom lavinntektsgrupper spiller mer enn høyinntektsgrupper, vil det være grunn til å hevde at skatteinnkreving gjennom pengespill er en regressiv skatt. Likevel kan vi ikke se på parameterestimatet β_1 fra likning (4.1) for å besvare spørsmål (2). Jeg ønsker snarere å estimere *korrelasjonen* mellom inntekt og pengespill, og kan *ikke* benytte den partielle effekten av inntekt på utgifter til pengespill.

Anta at vi formulerer regresjonsmodellen fra (4.1) med $n = 2$, det vil si at modellen sier at kun x_1 (inntekt) og x_2 (utdanning) påvirker y (spilleutgifter). Videre antar vi $\beta_1 = 0$, og $\beta_2 < 0$, samt $Cov(x_1, x_2) > 0$ for populasjonen. I dette tilfellet vil vi predikere at en rørlegger med høy inntekt spiller mer enn en akademiker med samme inntekt, men høyinntekts-rørleggeren vil spille like mye som en rørlegger med lav inntekt. Samtidig antar vi at utdanning er positivt korrelert med inntekt, slik at gjennomsnittsinntekten blant rørleggere er lavere enn akademikere. Da har vi $\beta_1 = 0$, men samtidig har vi at høyinntektsgrupper spiller mindre enn lavinntektsgrupper. Dette betyr ikke at de spiller mindre *fordi* de tjener mer; de spiller mindre fordi de har høyere utdanningsnivå, som har en negativ partiellem effekt på pengespillutgifter, og samtidig en positiv korrelasjon med inntekt, illustrert i figur 4.1.



Figur 4.1: Spuriøs sammenheng mellom inntekt og utgifter til pengespill

Til tross for at det ikke er noen kausalitet mellom inntekt og utgifter til pengespill i dette eksempelet, kan en regresjon av inntekt på pengespillutgifter være riktig metode for å fastslå hvorvidt pengespill kan sies å være en regressiv, proporsjonal eller progressiv skatt. Kun da vil jeg besvare spørsmål (2), som omhandler korrelasjonen mellom variablene inntekt og utgifter til pengespill. Jeg spør om det er konsumenter med høye eller lave inntekter som betaler for offentlige svømmehaller gjennom lotterideltakelse, men ikke *hvorfor* det er slik. Spørsmål (1) går inn på mulige *forklaringer* av en eventuell sammenheng mellom inntekt og lotteriutgifter. Det kan godt tenkes at jeg besvarer spørsmål (2) med at lavinntektsgrupper har høyere utgifter til pengespill, men at jeg ved å besvare spørsmål (1) viser at dette skyldes bakenforliggende variabler som utdanning, yrkesdeltakelse eller liknende.

Samtidig er det viktig å påpeke at parameterestimater fra multippel regresjon ikke skal tolkes som kausale sammenhenger i streng forstand; de måler kun partiell korrelasjon mellom forklaringsvariablene og spilleadferd. Eksempelvis kan vi anta at individuelle holdninger til risiko er med å bestemme hvorvidt en konsument deltar i pengespill. Samtidig er det rimelig å anta at holdning til risiko også påvirker øvrige forklaringsvariabler som utdanningsnivå og inntekt (via yrkesvalg). Vi har da en forklaringsvariabel som ender opp i feilleddet, men som er korrelert med de øvrige forklaringsvariablene og derfor gir forventningsskjevne estimater av ceteris paribus effekten. Kausalitet forutsetter også at vi kan bestemme effektens retning, eller hva som påvirker hva. For noen av våre forklaringsvariabler er retningen uproblematisk, vi har grunn til å anta at kjønn påvirker spilleadferd, men at spilleadferden sjelden fører til kjønnsskifte. Derimot er retningen mellom spilleadferd og arbeidsledighet mindre tydelig fra et teoretisk utgangspunkt; spiller konsumenten fordi hun er arbeidsledig, eller ble hun arbeidsledig fordi hun spilte?

For å si noe om den kausale sammenhengen mellom inntekt og utgifter til pengespill, kunne en løsning være å finne gode instrumentvariabler og gjøre en IV regresjon. En god instrumentvariabel er ikke korrelert med feilleddet, men korrelert med variabelen vi ønsker å estimere (Wooldridge 2009: 508). En annen løsning kan være å estimere med "fixed effects" på paneldata, dersom de utelatte variablene er konstante i tid. Dette kan være en problematisk forutsetning for variabler som måler holdning til risiko, siden slike holdninger kan antas å variere for ulike livsfaser.

4.1 Metodevalg i tidligere forskning

Inntektselastisiteten defineres som

$$(4.2) \quad \epsilon = \frac{\partial Q}{\partial M} \frac{M}{Q} = \frac{\partial \ln(Q)}{\partial \ln(M)}$$

hvor Q er etterspurt mengde av et gode (her pengespill), og M er inntekt. Det ser ut til å være en utbredt oppfatning blant økonomer og andre samfunnsforskere at inntektselastisiteten generelt kan gi svar på hvorvidt pengespill er en form for regressiv skattlegging. Dette er uproblematisk i tovariatt analyse som vi finner i Clotfelter og Cook (1987) og i første del av Beckert og Lutter (2009). Problemene oppstår når vi tar inn øvrige forklaringsvariabler på pengespillomsetning i analysen, slik vi finner i andre del av Beckert og Lutter (2009), samt Pérez og Humphreys (2011) og Worthington (2001). Her finner vi påstander som

- "A multivariate model offers a substantially more differentiated possibility for analyzing regressivity. A measure of regressivity can be obtained by estimating the income elasticity of the demand for lottery tickets." (Beckert og Lutter 2009: 481)
- "Negative or small positive income elasticities [implies] regressive implicit lottery taxes." (Pérez og Humphreys 2011: 566)
- "[A]n elasticity greater than unity would indicate progressivity (...)" (Worthington 2001: 332).

Disse tre studiene benytter multippel regresjon, og vil følgelig ende opp med inntektselastisiteter som ikke nødvendigvis forteller noe om graden av regressivitet eller progressivitet, dersom vi med disse begrepene sikter til hvordan skattebyrden fordeles mellom ulike inntektsgrupper i samfunnet.

4.2 Datamateriale

Pengespill er en aktivitet som kan være forbundet med skyldfølelse, spesielt for høye spillnivåer. Flere studier finner en tendens til underrapportering av utgifter til pengespill (Pérez og Humphreys 2011: 562). Vi har også grunn til å tro at storspillere har mindre kontroll med husholdningsutgiftene, og derfor gir et unøyaktig estimat av sine spilleutgifter. Et ideelt datasett for å undersøke problemstillingene som reises i denne oppgaven, vil være basert på observert konsumadferd snarere enn selvrappotering.

I Norge er Norsk Tipping operatør for pengespill som tilsvarer om lag 2/3 av nettoomsætningen i det regulerte spillmarkedet i Norge. Den siste tredjedelen består hovedsakelig av travsport med Norsk Rikstoto som operatør, samt bingospill (se tabell 2.1). Alle spillene til Norsk Tipping med unntak av skrapelodd²⁹ registreres på et unikt spillnummer, som igjen er knyttet til et personnummer av identifiseringstjenesten Buypass. Dermed kan nær 2/3 av det totale pengespillmarkedet i Norge i teorien knyttes opp til et personnummer, som igjen kan knyttes opp til eksempelvis inntekt, arbeidsstatus, alder, kjønn og bosted via offentlige registre fra skatteetaten og trygdeetaten. En institusjon med infrastruktur til å behandle sensitive persondata kan dermed bruke personnummer til å sette sammen et datasett med en rekke sosioøkonomiske forklaringsvariabler, for så å erstatte personnummeret med en anonym id for hver observasjon. Jeg har henvendt meg til Norsk Tipping for å undersøke mulighetene for at selskapet bidrar med sine kundeopplysninger i et slikt forskningsprosjekt, men foreløpig viser Norsk Tipping til sin kundeavtale hvor de garanterer at personopplysninger ikke gis ut til tredjepart.³⁰ Jeg må derfor benytte alternative datakilder for å besvare spørsmålene jeg reiser i denne oppgaven.

4.2.1 Omsetningstall, norske kommuner

Norsk Tipping publiserer omsetningstall for alle norske kommuner på sine nettsider. Jeg tar utgangspunkt i år 2009, og har satt sammen et datasett med alle norske kommuner (N=430) hvor omsetning på Norsk Tippings spill kombineres med kommunedata fra Statistisk Sentralbyrå. Jeg har hentet følgende variabler fra nettverktøyet til SSB, Statistikkbanken, for alle kommuner:

- Medianinntekt etter skatt. (Tabell 06944)
- Folkemengde (Tabell 05231)
- Antall kvinner (Tabell 07459)
- Andel bosatt i tettbebygd strøk (Tabell 05212)
- Registrert arbeidsledige (Tabell 06900)
- Antall med grunnskoleutdanning, samt antall med universitets- og høyskoleutdanning (Tabell 06983)
- Netto driftsutgifter til kultursektoren pr innbygger (Tabell 04909)

²⁹ Omsetning skrapelodd (Flax) 2009: 978 millioner kroner; anslagsvis 8 % av omsetningen til Norsk Tipping. (Norsk Tipping 2009)

³⁰ Tredjepart omfatter ikke Lotteritilsynet eller Kulturdepartementet, se Norsk Tippings kundeavtale.

- Utgifter til sosialstønad, pr innbygger (Tabell 05074)
- Antall alders- og uførepensjonister (Tabell 03384)
- Alderssammensetning (Tabell 07459)

Foruten forklaringsvariablene diskutert i kapittel 3, har jeg også tatt med driftsutgifter til kultur i kommunen, samt utgifter til sosialstønad og antall trygdemottakere. Jeg antar at driftsutgifter til kultursektoren er positivt korrelert med det generelle kulturtilbudet i kommunen, og at slike kulturgoder er substitutter for pengespill som underholdningsgode. Flere studier viser en sammenheng mellom høye utgifter til pengespill og sosiale problemer som rus, psykiske lidelser, lovbrudd og liknende (Fekjær 2002: 75). Jeg antar at utgifter til sosialstønad pr innbygger er en proxyvariabel som kan måle graden av slike sosiale problemer i kommunen. Som diskutert i kapittel 3, kan pengespill som bingo og spilleautomater være spesielt attraktive for konsumenter med mye fritid, siden disse spillene kan ha lave kostnader per tidsenhet. Jeg tar derfor med variabelen antall alders- og uførepensjonister for å se om det er en større spilleomsetning i kommuner med høye andel trygdemottakere.

Omsetningstall fra Norsk Tipping baserer seg på kommisjonærens beliggenhet, ikke spillerens hjemkommune. Vi kan altså ha en målefeil i den avhengige variabelen som kan påvirke variansen til parameterestimaterne, men forutsatt at målefeilen ikke er systematisk korrelert med de uavhengige variablene vil parameterestimaterne være forventingsrette (Wooldridge 2009: 314-316). Denne forutsetningen kan være problematisk. Regionsentra og byer kan ha stor arbeids- og handelstrafikk fra nabokommuner, og dette vil medføre større målefeil. Målefeilen vil da være korrelert med forklaringsvariablene som omhandler kommunestørrelse og andel innbyggere i tettbebygd strøk.

4.2.2 Forbruksundersøkelsen

I tillegg til aggregerte kommunedata vil jeg benytte individdata i en separat analyse. I utgangspunktet ønsker jeg å benytte husholdninger snarere enn enkeltindivider. En husholdning er i mange tilfeller en udelelig økonomisk enhet, og det kan være vanskelig å bestemme hvilket medlem som investerer i aksjefond eller deltar i pengespill.³¹ Statistisk Sentralbyrå gjennomfører årlig en forbruksundersøkelse, hvor et utvalg av norske husholdninger fører et detaljert regnskap over sitt forbruk. Forbruksundersøkelsen fra 2006 hadde et bruttoutvalg på 2166 husholdninger, med frafall på 1168 husholdninger. (54 %). Alle

³¹ Økonomifagets nære binding til *husholdningen* har som kjent også en etymologisk dimensjon.

husholdninger fører sitt forbruk i en 14 dagers periode, fordelt likt over hele året slik at 1/26 av respondentene fører de to første ukene i januar osv. Forbruksundersøkelsen fra 2005 hadde et nettoutvalg på 2162 husholdninger, med frafall på 910 husholdninger. (49 %). Jeg antar lønns- og prisstigning som ubetydelig over to år, og genererer et datasett (N=2089) av alle respondentene i 2005 og 2006.

Husholdningenes utgifter til pengespill registreres i Forbruksundersøkelsen under variablene *verva745*: Pengelotteri og tipping; *verva746*: Bingo; samt *verva747*: Andre lotteri- og spilleutgifter, herunder Flax og V6-Trav. Det er Statistisk Sentralbyrå som klassifiserer spilleutgiftene i de tre underkategoriene, basert på opplysningene i føringsheftet. Jeg antar at variabelen *verva745* dekker Norsk Tippings pengespill, med unntak av skrapeloddene.³² Variabelen *verva747* antar jeg skal dekke spilleautomater³³, hestesport, skrapelodd, øvrige lotterier samt spill på utenlandske nettsteder. I veiledningen som følger føringsheftet, instrueres husholdningen til at gaver og gevinster skal føres separat: "Gaver og gevinster du/dere mottar fra noen utenfor husholdningen, føres i et eget felt nederst på siden." (Holmøy 2008: Vedlegg 5, side 15). Dermed antar jeg at *verva745* er brutto utgifter til denne typen pengespill. (*Verva747* tolker jeg som delvis angitt på nettoform siden gjenspill av smågevinster på spilleautomater innenfor en og samme spilleperiode er vanlig.) Jeg velger å benytte *verva745*: Utgifter til pengelotteri og tipping som avhengig variabel.

Forbruksundersøkelsen operer med en *hovedperson*, som er den respondenten identifiserer ved spørsmålet "Hvem i husholdningen bidrar mest til økonomien?". Jeg velger å benytte hovedpersonens alder som en proxyvariabel for aldersnivået i husholdningen. Videre antar jeg at hovedpersonens utdanningsnivå uttrykker utdanningsnivået i husholdningen. Jeg vil kun se på hvorvidt hovedpersonen er i arbeid, framfor å differensiere mellom husholdninger hvor én av to står utenfor arbeidslivet og husholdninger hvor begge er i arbeid.³⁴ Siden det er husholdningen og ikke individer som er respondent, kan vi ikke si noe om individuelle egenskaper på bakgrunn av data fra Forbruksundersøkelsen. Derfor kan vi ikke fange opp effekten av kjønn på utgifter til pengespill. Merk at data på individnivå heller ikke ville gitt en optimal kjønnsvariabel, gitt at mange individer er medlem av en flerkjønnnet husholdning som velger sitt konsum i fellesskap.

³² Bekreftet av seniorrådgiver Eiliv Mørk ved SSB per e-post datert 02.04.2012.

³³ Spilleautomater ble ikke fjernet fra det norske spillemarkedet før sommeren 2007.

³⁴ Jeg er klar over at en husholdning også kan bestå av flere enn to personer i full jobb, men normalt består en norsk husholdning av en eller to forsørgere i tillegg til eventuelle barn/eldre.

Respondentene i Forbruksundersøkelsen 2005 og 2006 har i gjennomsnitt utgifter lik 1971 kroner på pengelotterier og tipping (*verva745*). Dersom vi antar 2,2 personer i hver husholdning og 4,64 millioner innbyggere i 2006, har vi totalt 2 109 090 husholdninger. Totale estimerte utgifter til pengelotterier og tipping fra FU blir da i overkant av 8,3 milliarder norske kroner. Til sammenlikning finner jeg at samlet omsetning for pengelotterier og tipping fra Norsk Tipping i 2005/2006 var 16,2 milliarder norske kroner.³⁵ Vi ser at det er et betydelig avvik mellom estimert totalomsetning på pengelotterier og tipping fra Forbruksundersøkelsen sammenliknet med omsetningsdata fra Norsk Tipping. Dette indikerer at vi har et problem med underrapportering i husholdningene, altså at vi har en målefeil i den avhengige variabelen som kan påvirke variansen til parameterestimatene. Det er grunn til å tro at denne målefeilen kan være systematisk korrelert med de uavhengige variablene, eksempelvis kan vi anta at husholdninger med lav inntekt har bedre oversikt over småutgifter enn husholdninger med høy inntekt. Vi vil da kunne få forventningsgale estimater, og resultatene bør derfor tolkes med forsiktighet. Samtidig er det verdt å minne om at datamaterialet fra Forbruksundersøkelsen benyttes til utregning av konsumprisindeksen, i nasjonalregnskapet og en rekke andre offentlige økonomiske analyser. På nåværende tidspunkt anser jeg datamaterialet fra FU som det beste alternativet til å undersøke konsum av pengespilltjenester i norske husholdninger.

Som beskrevet over er Forbruksundersøkelsen basert på føringshefter i husholdninger over en 14-dagers periode. Alt forbruk i perioden multipliseres med 365/14 for å angi årlig forbruk. Dersom variabelen *verva745*: Utgifter til pengelotteri og tipping tar verdien 0 betyr dette at husholdningen ikke har deltatt i Norsk Tipping sine spill i føringsperioden. For den enkelte husholdning vil *verva745* avvike fra faktiske utgifter til pengespill dersom husholdningen spiller sjelden. Eksempelvis vil husholdninger som kun kjøper 5-ukers lottokuponger i 3 av 5 tilfeller feiltolkes som ikke-spillere, og i 2 av 5 tilfeller ha et høyere estimert årlig forbruk enn faktisk forbruk.³⁶ Jay Stewart (2009) diskuterer utfordringene når vi har informasjon om individers aktiviteter over et kort tidsrom, slik at en stor andel av observasjonene har nullverdier selv om de deltar i aktiviteten på jevnlig basis. Dette gjelder også for forbruksdata: "(...) [E]xpenditures on clothing are often zero in expenditure diaries, but casual empiricism suggests that virtually everybody purchases clothing." (Stewart 2009: 2). I vårt datamateriale har 38,2 % av respondentene deltatt i pengespill, mens vi finner en høyere andel spillere i

³⁵ Totalomsetning 2006: 9619 millioner kroner, hvorav Flax 761 millioner kroner. Totalomsetning 2005: 9203 millioner kroner, hvorav Flax 883 millioner kroner. (Norsk Tipping 2006: 10)

³⁶ Merk at estimert samlet utgift for populasjonen blir korrekt, dersom utvalget er tilfredsstillende stort.

norske undersøkelser. Lund og Nordlund (2003: 47) finner at 81 % av befolkningen har deltatt siste 12 mnd. Øren og Bakken (2007: 21) rapporterer at 70 % har deltatt i pengespill siste år, av disse spiller 35 % sjeldnere enn månedlig. Dette indikerer at vi har en andel observasjoner med nullverdier i utvalget som vil ha $verva_{745} > 0$ hvis føringsperioden forlenges.

Grunnet lav utvalgsskjevhet veker jeg ikke datamaterialet fra Forbruksundersøkelsen. Dokumentasjonsrapporten som følger FU 2005 understøtter dette valget: "Hvis temaet en studerer (den avhengige variabelen) har en særlig sterk sammenheng med kjennetegn som er skjevt fordelt i utvalget, kan en vurdere å vekte datamaterialet for å minske effekten av skjevhetene. I de fleste tilfeller vil skjevhetene være for små til at dette har noen hensikt." (Holmøy 2008: 13)

4.3 Økonometrisk modell

4.3.1 Kommuner

Med utgangspunkt i datasettet for norske kommuner vil jeg gjøre en multippel regresjonsanalyse med den hensikt å forklare pengespillutgifter i kommunen. Konkret vil jeg estimere følgende økonometrisk modell;

$$(4.3) \quad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{12} x_{12} + u$$

der

- y er omsetning på Norsk Tippings pengespill per innbygger
- x_1 er medianinntekten etter skatt i kommunen
- x_2 er prosent registrert arbeidsledighet i kommunen
- x_3 er prosentandelen innbyggere med grunnskole som høyeste utdanning
- x_4 er prosentandelen innbyggere med universitet- eller høyskoleutdanning
- x_5 er prosentandelen innbyggere bosatt i tettbebygd strøk
- x_6 er prosentandelen menn bosatt i kommunen
- x_7 er netto driftsutgifter for kultursektoren i kommunen per innbygger
- x_8 er sosialstønadsutbetalinger pr innbygger, 16 år og oppover
- x_9 er prosentandelen alderspensjonister i kommunen
- x_{10} er prosentandelen uførepensjonister i kommunen

- x_{11} er prosentandelen innbyggere mellom 16 og 44 år
- x_{12} er en dummyvariabel som tar verdien 1 hvis kommunen ligger i Nordland, Troms eller Finnmark

Jeg vil være oppmerksom på høy grad av multikollinearitet, og eventuelt fjerne variabler fra regresjonsanalysen dersom vi har få signifikante t-verdier. Jeg vil forsøke ulike funksjonelle former, og teste for heteroskedastisitet som vi ofte finner i denne type tverrsnittdata.

Parameterestimaterne β_1, \dots, β_n kan tolkes som effekten av den aktuelle forklaringsvariabelen på omsetning av Norsk Tippings pengespill i norske kommuner. Den konkrete tolkningen om avhengig variabel og de uavhengige variable er på logaritme- eller lineær form. Jeg vil gi en presis tolkning av parameterestimaterne når funksjonell form er valgt.

Som diskutert innledningsvis i dette kapittelet, vil parameterestimatet β_1 ikke kunne besvare hvorvidt lavinntektskommuner i større grad enn høyinntektskommuner finansierer idretts-, kultur- og organisasjonsliv i Norge. Det kan tenkes at det spilles for like beløp i to kommuner hvor medianinntekten er ulik, men hvor alle andre variable holdes likt; altså at den partielle effekten av inntekt (*ceteris paribus*) er lik 0. Likevel er det den tovariate regresjonen av inntekt på spilleomsetning som vil besvare spørsmålet om hvorvidt omsetning i pengespill – som igjen uttrykker den implisitte skattebyrden – er proporsjonal, tiltakende eller fallende for økt inntektsnivå:

$$(4.3.1) \quad y = \theta_0 + \theta_1 x_1 + u$$

Hvor

- y er omsetning på Norsk Tippings pengespill per innbygger
- x_1 er medianinntekten etter skatt i kommunen

Et negativt estimat for θ_1 tilsier at omsetningen – og dermed skattebyrden – er høyere i kommuner med lav inntekt enn i kommuner med høy inntekt. Dersom vi setter regresjonsfunksjonen på logaritmeform, vil vi ha følgende tolkning av parameterestimatet for inntekt:

$$(4.3.2) \quad \ln(y) = \theta_0 + \theta_1 \ln(x_1) + u$$

- $\widehat{\theta}_1 > 1$: Progressiv skatt. Høyere skatteprosent for høyere inntektsnivåer.
- $\widehat{\theta}_1 = 1$: Flat skatt. Innbetalt skatt via pengespill øker proporsjonalt med inntekt.
- $1 > \widehat{\theta}_1 > 0$: Regressiv skatt. Lavere skatteprosent for høyere inntektsnivåer, men høyere innbetalt skatt – i absolutte tall – for høyere inntektsnivåer.
- $\widehat{\theta}_1 = 0$: Koppskatt. Innbetalt skatt via pengespill er uavhengig av inntekten.
- $\widehat{\theta}_1 < 0$: "Bunnskatt". Innbetalt skatt – i absolutte tall – øker for lavere inntekt.

I tillegg til å estimere regresjonsfunksjonen vil jeg lage visuelle plott for å se etter ikke-lineære mønstre for inntektsvariabelen.³⁷

4.3.2 Husholdninger

Med utgangspunkt i datasett fra Forbruksundersøkelsen 2005 og 2006, vil jeg gjøre en multipel regresjonsanalyse med utgifter til Norsk Tippings pengespill som den avhengige variabel. I motsetning til på kommunenivå, inneholder datasettet for husholdninger hjørneløsninger for variablene som beskriver utgifter til pengespill. 66 % av husholdningene har ikke registrert utgifter til noen former for pengespill og kun 25 % av utvalget har positive verdier for *verva745*: Utgifter til pengelotteri og tipping. Som diskutert kan det se ut til at vi har observasjoner med nullverdier grunnet underrapportering i husholdningene, men også grunnet kort føringsperiode relativt til innkjøpsfrekvens. Michael Keen (1986) viser at å benytte minste kvadraters metode på data med korte føringsintervaller kan føre til forventningsskjevhet (upward bias). I tillegg til lav innkjøpsfrekvens antar jeg at både underrapportering og normale hjørneløsninger som skyldes husholdningens preferanser³⁸ fører til nullverdier for *verva745* i utvalget.

Tobit-modellen kan være et alternativ når vi har hjørneløsning for en ikke-triviell andel av populasjonen, men kontinuitet for positive verdier (Wooldridge 2009: 587). Denne modellen uttrykker observerte spilleutgifter, y , med hjelp av en underliggende latent variabel y^* ;

$$(4.5) \quad y^* = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + u$$

$$(4.6) \quad y = \max(0, y^*)$$

³⁷ Det kan eksempelvis tenkes at lavinntektskommuner og høyinntektskommuner spiller lite, og at kommuner med middels inntekt spiller mye. Da vil det være aktuelt å ta med et andregradsledd for denne regresjonen.

³⁸ En husholdning kan ha hjørneløsning grunnet et prinsipielt valg som gjelder for alle inntekts- og prisnivåer, eksempelvis for tobakk eller pengespill. I tillegg kan vi ha normale hjørneløsninger hvor konsum kun kan anta positive eller nullverdier, slik at vi ikke alltid ser en atferdsendring når pris eller inntekt endres marginalt.

hvor x_1, \dots, x_n er forklaringsvariabler på pengespillutgifter. Vi vil få ML (maximum likelihood) estimater av β_1, \dots, β_n ved å maksimere log-likelihood-funksjonen ved en numerisk tilnærming i en økonometrisk programpakke.

En begrensning med tobit-modellen er at den ikke skiller mellom valgene $y = 0$ versus $y > 0$ og mengden av y gitt $y > 0$; som fører til at $\partial P(y > 0 | \mathbf{x}) / \partial x_j$ og $\partial E(y | \mathbf{x}, y > 0) / \partial x_j$ alltid har samme fortegn (Wooldridge 2002: 536). Dette trenger ikke nødvendigvis stemme med virkeligheten, eksempelvis finner Beckert og Lutter (2009) ingen statistisk signifikant partiellemffekt av inntekt på valget om å spille eller ikke spille, samtidig som de finner at pengespill er et normalt gode for spillergruppen; altså at parameterestimatet for inntekt her er positivt. Det kan derfor argumenteres for å bruke en økonometrisk modell som deler inn konsum av pengespill i to steg; først tar husholdningen valget om hvorvidt den vil delta i pengespill, for deretter å bestemme konsumnivå. Beckert og Lutter (2009) estimerer først en logit-modell for hele utvalget, og benytter deretter minste kvadraters metode på undergruppen som deltar i pengespill. De får parameterestimerer på det binære valget *spille / ikke spille* fra logit-modellen, og deretter OLS³⁹ estimater på utgifter til pengespill; gitt at husholdningen spiller. Hindermodeller⁴⁰ estimerer trinnene simultant, og kan gi ulike parameterestimerer for selve valget om å delta/ikke delta og konsumnivået gitt deltakelse. I tillegg kan vi beregne gjennomsnittlig partiellemffekt av en forklaringsvariabel på avhengig variabel.

Cragg-modellen (Cragg 1971) er en hindermodell som kombinerer en probit-modell og en modell med *truncation*. Modellen kan uttrykkes

$$(4.7) \quad f(w, y | \mathbf{x}_1, \mathbf{x}_2) =$$

$$\{1 - \Phi(\mathbf{x}_1 \boldsymbol{\gamma})\}^{1(w=0)} [\Phi(\mathbf{x}_1 \boldsymbol{\gamma}) (2\pi)^{-0.5} \sigma^{-1} \exp\{-(y - \mathbf{x}_2 \boldsymbol{\beta})^2 / 2\sigma^2\} / \Phi(\mathbf{x}_2 \boldsymbol{\beta} / \sigma)]^{1(w=1)}$$

hvor Φ er standard kumulativ normalfordeling og w er en binær indikator lik 1 hvis den avhengige variabelen y er positiv, og 0 ellers. Vi kan ha ulike variabler som innvirker på det binære og det kontinuerlige valget, derfor har vi to vektorer $\mathbf{x}_1, \mathbf{x}_2$ med tilhørende parametere $\boldsymbol{\gamma}, \boldsymbol{\beta}$. Merk at dersom $\mathbf{x}_1 = \mathbf{x}_2$ og $\boldsymbol{\gamma} = \boldsymbol{\beta} / \sigma$ reduseres modellen til en Tobit-modell (Burke 2009). Jeg vil benytte kommandoen `craggit` i Stata, se Burke (2009) for utfyllende dokumentasjon. I korte trekk estimerer `craggit` ML estimater av $\boldsymbol{\gamma}, \boldsymbol{\beta}, \sigma$ simultant, og gir

³⁹ *Ordinary least squares*; eller minste kvadraters metode (MKM).

⁴⁰ "Hurdle models" (Wooldridge 2009: 595).

muligheter for enklere analyse av resultatene sammenliknet med å benytte **probit** og **truncreg** separat. Jeg vil estimere modellen med variablene

- w : Spiller. 1 hvis spilt siste 14 dager, ellers 0.
- y : verva745: Utgifter til pengelotteri og tipping.
- $x_{11} = x_{21}$: Husholdningens inntekt.
- $x_{12} = x_{22}$: Arbeidsledighet. 1 hvis husholdningen ikke har arbeidsinntekt, ellers 0.
- $x_{13} = x_{23}$: Grunnskoleutdanning. 1 hvis husholdningen har grunnskoleutdanning som høyeste utdanningsnivå, ellers 0.
- $x_{14} = x_{24}$: Høyere utdanning. 1 hvis husholdningen har høyskole-/universitetsutdanning, ellers 0.
- $x_{15} = x_{25}$: Enslig. 1 hvis enslig eller eneforsørger, ellers 0.
- $x_{16} = x_{26}$: Hovedpersonens alder.
- $x_{17} = x_{27}$: Hovedpersonens alder, kvadrert.

Disse forklaringsvariablene samsvarer med Beckert og Lutter (2009) og Pérez og Humphreys (2011).⁴¹ Merk at vi har $\mathbf{x}_1 = \mathbf{x}_2$; jeg antar at alle forklaringsvariablene kan ha innvirkning i begge trinn. Eksempelvis kunne en variabel som målte religiøsitet kun inngå i første trinn, siden jeg antar at dette kun innvirker på det prinsipielle valget om å spille eller ikke spille. Dessverre finner jeg ingen variabel i Forbruksundersøkelsen som kan måle graden av religiøsitet. Som diskutert i kapittel 3 er det grunn til å tro at norske menn spiller mer enn norske kvinner, men siden data er på husholdningsnivå kan jeg ikke fange opp en kjønnseffekt med datamaterialet fra husholdningene. Dersom jeg finner signifikante geografiske forskjeller på kommunenivå vil jeg legge til geografiske variabler også på husholdningsnivå. Merk at jeg har påvist målefeil i avhengig variabel, siden estimerte spilleutgifter basert på Forbruksundersøkelsen er betydelig lavere enn faktiske omsetningstall for Norsk Tipping. Målefeilen kan være korrelert med de avhengige variablene, og kan derfor føre til inkonsistente parameterestimer.

Jeg vil også se på korrelasjonen mellom utgifter til pengespill og husholdningenes inntekt for å undersøke om implisitte skatter via pengespill rammer ulikt i forskjellige inntektssegmenter

⁴¹ Beckert og Lutter har i tillegg med kjønn, statsborgerskap og bosted i by. Ingen av parameterestimatene er signifikant forskjellige fra null. Pérez og Humphreys har med kjønn, og finner signifikant positiv effekt av variabelen mann.

av befolkningen. Siden en rekke observasjoner har nullverdier for *verva745*, kan jeg ikke estimere en regresjonsfunksjon på logaritmeform. Samtidig antar jeg at det kan være en ikke-lineær sammenheng mellom inntekt og utgifter til pengespill. Jeg velger derfor å dele inn utvalget i inntektskvintiler, og uttrykker andel av inntekten som går til pengespill t for hvert kvintil i som

$$(4.8) \quad t_i = \frac{1}{H} \sum_{h=1}^H \frac{verva745_h}{inntekt_h}, H = N/5 \quad i = 1, \dots, 5$$

hvor h indikerer den enkelte husholdning, $H = N/5$ er antall husholdninger i hvert kvintil, totalt fem kvintiler. Dette samsvarer med metoden i Beckert og Lutter (2009), og jeg har dermed et sammenlikningsgrunnlag fra det tyske statslotteriet.

5 Resultater og diskusjon

5.1 Resultater fra multippel regresjon, kommuner

Jeg starter med en modell som er lineær i avhengig og uavhengige variable, og tester for heteroskedastisitet med en Breusch-Pagan test. Testen gir en kji-kvadrat verdi på 17,4 og jeg forkaster H_0 : konstant varians i feilleddene. Dette kan skyldes heteroskedastisitet i datamaterialet, men kan også være en indikasjon på at jeg benytter gal funksjonell form. Jeg plotter medianinntekt mot omsetning i kommunene og ser at vi bør sette inntekt og omsetning på logaritmeform. Jeg estimerer regresjonslinjen

$$(5.1) \quad \ln(y) = \beta_0 + \beta_1 \ln(x_1) + \beta_2 x_2 + \dots + \beta_{11} x_{11} + u$$

og får parameterestimerer gjengitt i kolonne (1) i tabell 5.1.

Jeg tester for heteroskedastisitet med en Breusch-Pagan test, og kjører også Ramseys RESET test. Testene gir ikke indikasjoner på feilspesifikasjon eller heteroskedastisitet. Jeg tester også for grad av multikollinearitet og finner tilfredsstillende verdier.

Jeg finner ingen alderseffekt på pengespillomsetning med kategoriene 18 til 44 år, 44 til pensjonistalder, og pensjonister. Jeg finner heller ingen effekt av andelen uføre i kommunen. Kjønnfordelingen i norske kommuner er relativt lik⁴², så variabelen x_5 fanger ikke opp en eventuell kjønnseffekt på pengespillomsetning. Jeg finner videre at andelen innbyggere med høyere utdanning heller ikke har en effekt.

Parameterestimatet for $\ln(x_1)$ – den naturlige logaritmen av inntekt – vil på husholdnings/individnivå tolkes som inntektselastisiteten for Norsk Tippings pengespill. På kommunenivå har vi medianinntekt som uavhengig variabel, og kan tolke β_1 som margineffekten av logaritmen av inntekt på logaritmen av omsetning. For en økning i medianinntekt med 1 %, gitt at øvrige variabler holdes konstante, estimerer jeg tilnærmet 0,72 % lavere omsetning på Norsk Tippings pengespill i kommunen.⁴³ Jeg finner altså en negativ inntektseffekt i norske kommuner som både er økonomisk og statistisk signifikant når vi kontrollerer for andre typiske forklaringsvariabler på pengespill.

⁴² 406 av 430 kommuner har andelen menn i intervallet 48 % - 52 %.

⁴³ Jeg kan regne dette ut eksakt med uttrykket $(1,01^{\beta_1} - 1) = (1,01^{-0,72} - 1) = (0,992861 - 1) = -0,007139$, det vil si at for en økning i medianinntekt med 1 % estimerer modellen 0,714 % lavere omsetning.

Tabell 5.1: Regresjonsmodell, spilleomsetning i norske kommuner.

VARIABLER	(5.1) ln(Omsetning, Norsk Tipping)	(5.2) Ln(Omsetning, Norsk Tipping)
ln(x₁): ln(Medianinntekt, kroner)	-0,720*** (0,222)	-1,411*** (0,121)
x₂: Arbeidsledige, prosent	0,0424** (0,0175)	
x₃: Innbyggere med grunnskoleutdanning, prosent	0,0150*** (0,00471)	
x₄: Innbyggere med høyere utdanning, prosent	0,00771 (0,00505)	
x₅: Menn i kommunen, prosent	-0,00968 (0,0133)	
x₆: Utgifter til kultur per innbygger, kroner	3,39e-05*** (8,01e-06)	
x₇: Utgifter til sosialstønad per innbygger, kroner	6,26e-05* (3,42e-05)	
x₈: Alderspensjonister, prosent	0,00797 (0,0121)	
x₉: Uførepensjonister, prosent	0,00168 (0,00982)	
x₁₀: Unge voksne – 18-44 år – prosent.	0,00499 (0,00991)	
x₁₁: Sørlandskommune	-0,243*** (0,0544)	
Konstant	16,48*** (3,286)	25,94*** (1,565)
N	430	430
R-kvadrat	0,367	0,240
Justert R-kvadrat	0,350	0,238

Standardfeil i parenteser. *** p<0,01, ** p<0,05, * p<0,1

Parameterestimatet for x_2 viser at andelen arbeidsledige har en statistisk signifikant effekt på spilleomsetning i norske kommuner. Det er høyere pengespillomsetning i kommuner med høy ledighet; dersom ledigheten stiger med ett prosentpoeng og øvrige forklaringsvariabler holdes konstant, estimerer jeg en økt pengespillomsetning på 4,24 %.⁴⁴ Ved første øyekast kan dette se lite ut, siden vi normalt finner 1 – 5 % arbeidsledighet i norske kommuner; kun 6 kommuner har høyere ledighet enn 5 %. Samtidig er det verdt å bemerke at parameterestimatet ikke måler den kausale effekten av arbeidsledighet på pengespill, her bør vi snarere anta at arbeidsledigheten i en kommune korrelerer med andre utelatte forklaringsvariabler. Dersom jeg mot bedre viten forutsetter at estimatet for arbeidsledighet

⁴⁴ Se Wooldridge (2009: 46) for en oversikt over tolkning av parameterestimerer i en modell som kombinerer log-level og log-log.

måler den faktiske *ceteris paribus* effekten på pengespillomsetning, det vil si at hvert prosentpoeng arbeidsledighet fører til 4,24 % økning i *pengespillomsetning per innbygger*, har de nye arbeidsledige økt den totale pengespillomsetningen i kommunen med et urimelig høyt beløp tatt gruppas størrelse i betraktning. Eksempelvis kan vi anta denne kausale sammenhengen i en kommune med 200 innbyggere, hvor det spilles for 400 000 kroner per år, det vil si 2000 kr per innbygger per år. Vi antar at halvparten, det vil si 100 personer, er i arbeidsstyrken. Dersom én innbygger blir ledig, det vil si at arbeidsledigheten øker med ett prosentpoeng, vil dette øke omsetningen til $2000 * 1,0424 = 2084,80$ kroner per innbygger, totalt 416 960 kr for hele kommunen. Dersom øvrige innbyggere holder sitt pengespill konstant, spiller den arbeidsledige nå for 16 960 kroner mer per år enn da hun var i arbeid, det vil si en økning på drøye 800 %. Dette illustrerer urimeligheten av å tolke parameterestimater som kausale sammenhenger.

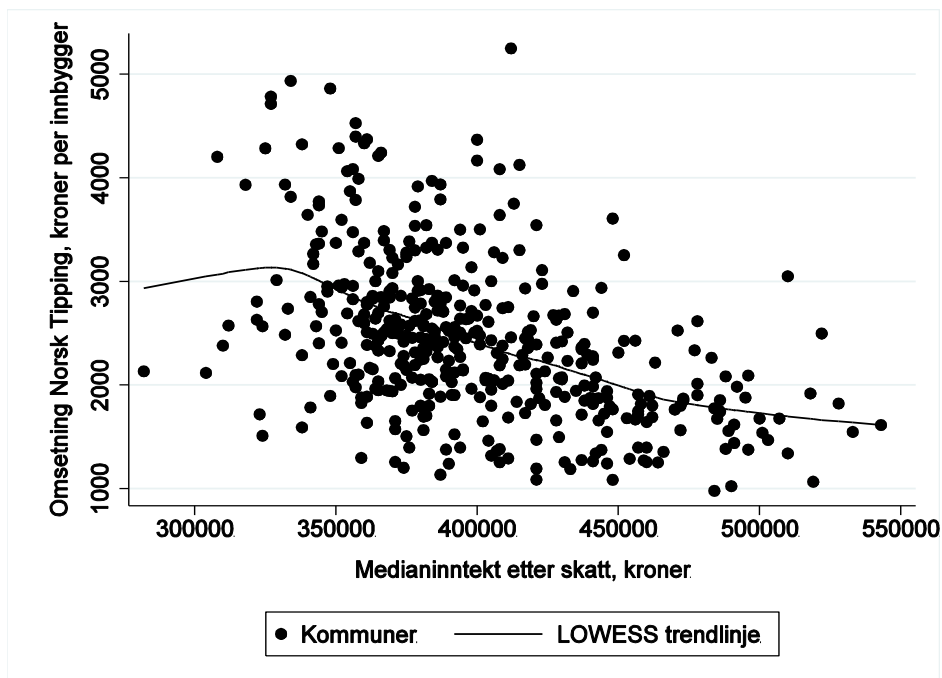
Jeg finner en effekt av andelen innbygger med grunnskoleutdanning som høyeste utdanning som både er statistisk og økonomisk signifikant; dersom denne andelen stiger med ett prosentpoeng forventer vi en økning i pengespill per innbygger på 1,5 %. Vi ser altså at lavt utdanningsnivå i kommunen korrelerer positivt med omfanget av pengespill, korrigert for øvrige forklaringsvariabler. Kulturutgifter per innbygger i kommunen har en uventet positiv effekt på spilleomsetning. I kapittel 3 antar jeg at øvrige kulturtilbud i kommunen er substitutter for pengespill. Estimatet for β_6 kan indikere at dette ser ikke ut til å stemme for Norsk Tippings spill, jeg finner snarere at høye kulturutgifter er positivt korrelert med spilleomsetning.⁴⁵

Jeg har forsøkt ulike geografiske variabler som kan ha effekt på pengespillomsetning. Innledningsvis forsøkte jeg å dele inn kommunene etter landsdel, og konstruerte dummyvariabler for Nord-Norge, Trøndelag, Sørlandet og Vestlandet. Østlandet stod da igjen som siste kategori i konstantleddet. Jeg fikk signifikante parameterestimater, men også spesifikasjonsfeil i modellen. Videre forsøkte jeg med andelen bosatt i tettbebygde strøk samt en dummyvariabel for større byer, som også resulterer i spesifikasjonsfeil. Dersom jeg kun beholder Sørlandet som geografisk variabel, får jeg et signifikant parameterestimat og har

⁴⁵ Sosiologene Tone Knudsen Haarr og Anne Krogstad (2011) diskuterer begrepet "kulturelite", og siterer en aviskommentar hvor man kunne lese at kultureliten "har et sugerør ned i kommunekassen, og på toppen av det hele bruker penger som noen fulle sjøfolk". (Fædrelandsvennen 03.10.2005 i Haarr og Krogstad 2011: 22). Selv om Haarr og Krogstad i første rekke *dekonstruerer* kulturelitebegrepet i sin artikkel, viser de til en diskurs i den norske offentligheten hvor man forstår kultureliten som en utgiftspost på kommunebudsjettene. Dette finner jeg ikke støtte for i estimatet for β_6 , her ser det snarere ut til at fransk nybølgefilm og fotballtipping går hånd i hånd.

ingen spesifikasjonsproblemer. Fremgangsmåten beveger seg i retning av datagruving – *data mining* – og er problematisk da dette bryter med forutsetningene for minste kvadraters metode (Wooldridge 2009: 678). Jeg velger likevel å ta med dummyvariabelen Sørlandet og lar den fungere som en proxy for det norske "bibelbeltet". Ikke overraskende finner jeg en negativ effekt av denne variabelen på pengespillomsetning; $(e^{-0,243} - 1) = -0,216$, det vil si at sørlandskommuner har 21,6 % lavere omsetning enn resten av landet hvis øvrige variabler holdes konstant.

5.2 Pengespill som implisitt skatt, kommunenivå



Figur 5.1: Medianinntekt og spilleomsetning, norske kommuner

Jeg ønsker også å se på korrelasjonen mellom medianinntekt og spilleomsetning i norske kommuner, og estimerer modellen

$$(5.2) \quad \ln(y) = \theta_0 + \theta_1 \ln(x_1) + u$$

hvor y er omsetning på pengespill fra Norsk Tipping i kommunen og x_1 er medianinntekten i kommunen. Vi får et estimat for θ_1 gjengitt i andre kolonne i tabell 5.1. Jeg finner et negativt parameterestimat, $\widehat{\theta}_1 = -1,41$ som tilsier at det er høyere spilleomsetning i kommuner med lav medianinntekt. Dette understøttes av figur 5.1, som indikerer at dette er et mønster for hele inntektsintervallet, med et mulig unntak for kommunene med lavest medianinntekt.⁴⁶

Det negative estimatet for θ_1 indikerer at statlige pengespill er en regressiv implisitt skatt på kommunenivå; innbyggerne i lavinntektskommuner bidrar mer til Norsk Tippings overskuddsmottakere ikke bare relativt til inntektsnivået, men også i absolutte verdier.

5.3 Resultater fra multippel regresjon, husholdninger

Jeg estimerer Cragg-modellen (Cragg 1971)

$$(5.2) \quad f(w, y | \mathbf{x}_1, \mathbf{x}_2) =$$

$$\{1 - \Phi(\mathbf{x}_1 \boldsymbol{\gamma})\}^{1(w=0)} [\Phi(\mathbf{x}_1 \boldsymbol{\gamma}) (2\pi)^{-0.5} \sigma^{-1} \exp\{-(y - \mathbf{x}_2 \boldsymbol{\beta})^2 / 2\sigma^2\} / \Phi(\mathbf{x}_2 \boldsymbol{\beta} / \sigma)]^{1(w=1)}$$

med variablene $\mathbf{x}_1, \mathbf{x}_2$ som presentert i 4.3.2. I tillegg tar jeg med variabelen x_{18} : *Bosatt i Agder eller Rogaland*, motivert fra funnene i modellen på kommunenivå. Jeg antar at denne variabelen er en proxy for kulturell religiøsitet⁴⁷ som kun er korrelert med det prinsipielle valget om å spille eller ikke spille. Estimaten for $\boldsymbol{\gamma}, \boldsymbol{\beta}$ er presentert i tabell 5.2.

Med utgangspunkt i estimatene for $\boldsymbol{\gamma}, \boldsymbol{\beta}, \sigma$ kan vi beregne partielleffekter av alle uavhengig variabler, se Cragg (1971) og Burke (2009) for utfyllende dokumentasjon. For en gitt husholdning, er partielleffekten av en uavhengig variabel x_j for sannsynligheten for at husholdningen deltar i pengelotterier gitt ved

$$(5.3) \quad \frac{\partial P(y > 0 | \mathbf{x}_1)}{\partial x_j} = \gamma_j \phi(\mathbf{x}_1 \boldsymbol{\gamma})$$

hvor ϕ er standard normalfordeling.

⁴⁶ Dette kan også være statistisk støy. Dersom jeg dropper kommunen i datasettet med lavest medianinntekt (Verran kommune i Nord-Trøndelag), er trendlinjen negativ for hele det gjenværende utvalget.

⁴⁷ Med kulturell religiøsitet mener jeg at også sekulære husholdninger bosatt i "bibelbeltet" vil påvirkes av en religiøst motivert motstand mot pengespill.

Tabell 5.2: Regresjonsmodell, spilleomsetning norske husholdninger.

VARIABLER	Tier1 (γ) w: =1 hvis deltatt i pengespill siste 14 dager	Tier2 (β) y: Utgifter til pengelotterier og tipping
x₁: Husholdningens inntekt etter skatt	-1,67e-07* (9,39e-08)	0,033 (0,037)
x₂: Arbeidsledighet, =1 hvis ingen arbeidstakere	-0,920 (0,106)	25556,49 37688,78
x₃: Grunnskoleutdanning, =1 hvis høyeste utdanning	0,053 (0,080)	-13637,1 (24508,54)
x₄: Høyere utdanning, =1 hvis gjennomført høyskole- eller universitetsutdanning	-0,203*** (0,068)	-51426,12 (61561,32)
x₅: Enslig, =1 hvis eneforsørger eller enslig	-0,120 (0,082)	-28478,55 (40056,13)
x₆: Alder, hovedperson i husholdningen	0,071*** (0,016)	8022,954 (10486)
x₇: Alder kvadrert	-0,000567*** (0,000167)	-65,723 (90,016)
x₈: Agder og Rogaland, =1 hvis bosatt i Agder eller Rogaland	-0,156* (0,085)	-
Konstant	-2,335*** (0,384)	-401106 (498353.2)
Sigma(MLE av σ)	34985.1* (20951.23)	
N	2087	
Log likelihood	-7301,8313	

Standardfeil i parenteser. *** p<0,01, ** p<0,05, * p<0,1

Videre kan vi uttrykke partielleffekten av x_j på y : *Utgifter til pengelotterier*, gitt $y > 0$ som

$$(5.4) \quad \frac{\partial E(y|y > 0, \mathbf{x}_2)}{\partial x_j} = \beta_j [1 - \lambda(\mathbf{x}_2 \boldsymbol{\beta} / \sigma) \{ \mathbf{x}_2 \boldsymbol{\beta} / \sigma + \lambda(\mathbf{x}_2 \boldsymbol{\beta} / \sigma) \}]$$

Avslutningsvis uttrykker vi partielleffekten av x_j på den ubetingete forventningsverdien til y som

$$(5.5) \quad \frac{\partial E(y|\mathbf{x}_1, \mathbf{x}_2)}{\partial x_j} = \gamma_j \phi(\mathbf{x}_1 \boldsymbol{\gamma}) \times \{ (\mathbf{x}_2 \boldsymbol{\beta} + \sigma \times \lambda(\mathbf{x}_2 \boldsymbol{\beta} / \sigma)) \} \\ + \phi(\mathbf{x}_1 \boldsymbol{\gamma}) \times \beta_j [1 - \lambda(\mathbf{x}_2 \boldsymbol{\beta} / \sigma) \{ \mathbf{x}_2 \boldsymbol{\beta} / \sigma + \lambda(\mathbf{x}_2 \boldsymbol{\beta} / \sigma) \}]$$

Jeg beregner partielleffekt av x_1 (inntekt) på y (pengespillutgifter) i Stata for alle observasjonene, og finner APE⁴⁸ – gjennomsnittlig partielleffekt – som middelverdien av beregnede verdier (Wooldridge 2009: 583). Disse er presentert i tabell 5.3. Jeg kan ikke benytte standardfeil fra de predikerte partielleffektene som standardfeil for APE, og benytter derfor bootstrapping for å få standardfeil og p-verdier for mitt estimat (Burke 2009: 590).

Tabell 5.3: APE – gjennomsnittlig partielleffekter av x_1 : inntekt

	Koeffisient	Standardfeil*	p-verdi*
$\overline{\partial P(y > 0 x_1) / \partial x_1}$	-5.56e-08	4,57e-08	0,224
$\overline{\partial E(y y > 0, x_2) / \partial x_1}$	0,0010674	0,0005582	0,056
$\overline{\partial E(y x_1, x_2) / \partial x_1}$	-0,0000264	0,0002622	0,920

*Fra bootstrapping

Vi ser at gjennomsnittlig effekt av inntekt på det binære valget om å spille eller ikke spille er negativ, men ikke signifikant forskjellig fra null. Gjennomsnittseffekten av inntekt gitt at husholdningen spiller er positiv og signifikant på 0,10 nivå. Disse trekker altså i motsatt retning på den ubetingete effekten av inntekt på beløp brukt til pengespill, som ender opp med et estimat ikke signifikant forskjellig fra null.

I tillegg til statistisk signifikans er det viktig å diskutere den økonomiske signifikansen av våre forklaringsvariabler, samt hvor godt modellen passer til datamaterialet. Jeg finner at det er en positiv gjennomsnittseffekt av inntekt på spillebeløp i husholdningen, betinget av at husholdningen spiller. Estimaten for $\overline{\partial E(y | y > 0, x_2) / \partial x_1}$ (0,0010674) kan tolkes som at vi estimerer 10,70 i økt spillebeløp per år gitt inntektsforskjell mellom to husholdninger på 10 000 kr/år, og øvrige variabler holdt konstant. Dersom vi sammenlikner en spillende lavinntektshusholdning med inntekt etter skatt lik 200 000 kroner og en spillende høyinntektshusholdning med inntekt etter skatt lik 800 000 kroner⁴⁹, vil vi estimere at husholdningen med høy inntekt spiller for 640 kroner mer per år; øvrige variabler holdt konstant. Vi ser at effekten er relativt beskjeden, selv for ytterpunktene i vårt datamateriale.

⁴⁸ *Average Partial Effect*. Jeg benytter overstrek på likning (5.3), (5.4) og (5.5) i tabell 5.3 som betegnelse for middelverdien.

⁴⁹ 9 % av utvalget har $x_1 < 200\,000$, 8 % av utvalget har $x_1 > 800\,000$.

Videre ser jeg av parameterestimaterne i tabell 5.2 at det er kun Tier1 som gir statistisk signifikante parametere. Samtidig gir dette første steg i modellen en lav verdi for McFaddens pseudo R-kvadrat; uttrykt som $(1 - \mathcal{L}_{ur} / \mathcal{L}_o)$ hvor \mathcal{L}_{ur} er log-likelihood funksjonen til vår estimerte modell og \mathcal{L}_o er log-likelihood funksjonen til modellen med kun konstantledd (Wooldridge 2002: 465). Vi ser at når funksjonene nærmer seg hverandre, vil pseudo R-kvadrat nærme seg 0. I første trinn av denne regresjonen har jeg $(1 - \mathcal{L}_{ur} / \mathcal{L}_o) = 0,0342$ som må sies å være relativt lavt, og modellen forklarer derfor lite av variasjonen i observasjonene.

5.4 Pengespill som implisitt skatt, husholdningsnivå

I analysen av norske kommuner benyttet jeg en tovariater regresjon med variablene medianinntekt og spilleomsetning på logaritmeform for å undersøke graden av regressivitet i Norsk Tippings pengespill. Som diskutert i 4.2.2 er ikke dette mulig når vi har en stor andel observasjoner med nullverdier. Jeg sorterer observasjonene etter husholdningsinntekt, og setter fem inntektskvintiler i hvor andel av inntekten som går til pengespill for hvert kvintil uttrykkes som

$$(5.6) \quad t_i = \frac{1}{H} \sum_{h=1}^H \frac{verva745_h}{inntekt_h}, H = N/5 \quad i = 1, \dots, 5$$

Tabell 5.4: Totale utgifter til *verva745*: Pengelotterier og tipping, sortert etter inntektskvintil.

Inntektskvintil, %	Inntekt per hush, kr	Spilleutgifter per hush, kr	Relative spilleutgifter, %
0-20	194870	1621	0,83 %
20-40	351205	2469	0,70 %
40-60	469837	1954	0,42 %
60-80	568690	1772	0,31 %
80-100	938487	2050	0,22 %
Sum	504618	1973	0,39 %

Det fremgår av tabell 5.4 at vi finner ulike utgifter til Norsk Tippings pengespill i ulike inntektskvintiler. Husholdningenes spilleutgifter i absolutt beløp har likhetstrekk med funn på kommunenivå; det ser ut til at utgifter til pengespill initialt stiger med inntekt for deretter å falle, se figur 5.1. Samtidig må det bemerkes at forskjellen mellom gruppene absolutte spilleutgifter ikke er statistisk signifikant på 5 % nivå. Enveis variansanalyse fra Stata beregner en ikke-signifikant F-verdi⁵⁰, altså kan vi ikke utelukke at variasjonen mellom gruppene er tilfeldig. Videre finner jeg av Bonferronis sammenligningstest⁵¹ at variasjonen i all hovedsak er mellom første og andre inntektskvintil.

Selv om vi ikke kan utelukke at spillebeløpet er konstant i absolutte tall, ser vi at spillebeløp relativt til husholdningens inntekt er fallende med økte verdier for inntektskvintil. Dette samsvarer med funn fra lotterier i andre land (Beckert og Lutter 2009, Clotfelter og Cook 1987). Merk at spilleutgiftene i tabell 5.4 er basert på selvrapporing, og omfatter anslagsvis halvparten av Norsk Tippings omsetning i de aktuelle årene 2005-2006. Jeg antar at Norsk Tipping hadde 60 % av det norske spillmarkedet i 2006⁵², og totalutgifter på alle typer pengespill for norske husholdninger blir da om lag 3 ganger høyere enn oppgitt i tabell 5.4.

Det kan også være interessant å se hvordan spilleutgiftene er fordelt mellom ulike spillekvintiler, altså undersøke totale pengespillutgifter for ulike spillnivåer. Tidligere forskning har vist at en stor andel av total spilleomsetning kommer fra en liten andel av spillerne. Beckert og Lutter (2009: 480) finner at 25 % av spillerne står for 70 % av omsetningen i det tyske statslotteriet, mens Lund og Nordlund (2003: 8) estimerer at 10 % av spillerne står for 54 % av totale pengespillutgifter i Norge. Datamaterialet fra Forbruksundersøkelsen viser også at utgifter til pengespill er skjevt fordelt blant husholdningene. Tabell 5.5 sorterer spillende husholdninger etter utgifter til pengespill, og viser samlet beløp for hver gruppe. Som diskutert i avsnitt 4.2.2 har kun 25 % av utvalget positive verdier for *verva745*, og vi har da at 6,25 % av det totale utvalget, inkludert ikke-spillere, står for 59 % av den totale omsetningen på Norsk Tippings pengespill.

⁵⁰ $F = 1,96$. Med 5 grupper og 2087 observasjoner har vi $p = 0,098$.

⁵¹ *Bonferroni multiple-comparison test*

⁵² Se tabell 2.1

Tabell 5.5: Utgifter til *verva745: Pengelotterier og tipping*, sortert etter spillekvintil

Spillekvintiler, %	Spilleutgifter per hush, kr	Spilleutg, kr	Andel, %
0-25	1294	200564	4,89 %
25-50	3175	492117	11,99 %
50-75	6417	994627	24,24 %
75-100	15584	2415455	58,87 %
Sum (N=621)	1973	4102763	100,00 %

5.5 Suits-indeksen

Daniel Suits (1977) kalkulerte graden av progressivitet for ulike skatter med utgangspunkt i en Lorenz-kurve, hvor Lorenz-kurven representerer den akkumulerte andelen av skattebyrden for en gitt inntektsandel. Suits-indeksen S uttrykkes som

$$(5.7) \quad S = 1 - 2 \int_0^1 T(y) dy$$

hvor $T(y)$ er nettopp Lorenz-kurven. I dette tilfelle med diskrete verdier gjør jeg følgende tilnærming inspirert av Suits (1977):

$$(5.8) \quad \int_0^1 T(y) dy \approx \sum_{i=1}^5 \frac{1}{2} [T(y_i) + T(y_{i-1})] (y_i - y_{i-1})$$

Jeg beregner kumulerte inntekter og spilleutgifter for fem inntektskvintiler, og får verdiene i tabell 5.6. Merk at selv om skattebeløpet kun antas å være en gitt andel av kumulerte spilleutgifter (se avsnitt 3.4.1), vil andelen være lik for alle inntektskvintiler. Jeg setter likning (5.8) inn i likning (5.7) og beregner med verdier fra tabell 5.6:

$$(5.9) \quad S \approx 1 - 2 \sum_{i=1}^5 \frac{1}{2} [T(y_i) + T(y_{i-1})] (y_i - y_{i-1}) = -0,268$$

Negative verdier for Suits-indeksen indikerer en regressiv skatt. Beckert og Lutter (2009: 481) finner $S = -0,23$ for det tyske statslotteriet, og konkluderer med at dette indikerer en "(...)distinctly regressive distribution". Videre oppsummerer Beckert og Lutter funn fra 12

tidligere studier, som alle finner en negativ Suits-indeks for pengespill. Våre funn fra Forbruksundersøkelsen ser altså ut til å gi støtte til hypotesen om pengespill som en form for regressiv skattlegging.

Tabell 5.6: Kumulative husholdningsinntekter og utgifter til *verva745: Pengelotterier og tipping.*

Inntektskvintil	Kumulerte inntekter (y)	Kumulerte spilleutgifter [$T(y)$]
1	0,077	0,164
2	0,216	0,415
3	0,403	0,613
4	0,628	0,792
5	1	1

6 Avsluttende bemerkninger

6.1 Hovedfunn, kommuner

Med utgangspunkt i aggregerte data fra norske kommuner, har jeg forsøkt å besvare spørsmål (1):

Hvilke sosioøkonomiske variabler kan forklare forskjeller i utgifter til pengespill hos norske konsumenter?

Multippel regresjon av medianinntekt og øvrige forklaringsvariabler på spilleomsetning gir negativt parameterestimat for medianinntekt. Videre estimerer jeg høyere spilleomsetning i kommuner med høy arbeidsledighet, og høyere omsetning i kommuner hvor det er en stor andel med grunnskole som høyeste fullførte utdanning. Sørlandskommuner spiller mindre enn resten av landet, og det er en positiv partiellkorrelasjon mellom kulturutgifter per innbygger og spilleomsetning. Negativt parameterestimat for medianinntekt kan indikere at pengespill er et mindreverdige gode, hvor etterspørselen synker for økt inntekt. Samtidig må det bemerkes at selv om medianinntekten uttrykker inntekten til en typisk husholdning i kommunen, sier den ikke nødvendigvis noe om hvordan inntektsnivået varierer for hele inntektsintervallet.

Videre forsøker jeg å besvare spørsmål (2):

Flere fellesgoder finansieres av pengespill. Bryter dette med prinsipper om progressive eller proporsjonale skatter?

En tovariabel regresjon av medianinntekt på spilleomsetning i kommunene gir et negativt parameterestimat for inntekt. Dette understøttes av visuelle plott, som viser en klar negativ sammenheng mellom medianinntekt og pengespillomsetning. Innbyggerne i kommuner med lav medianinntekt bidrar i større grad til Norsk Tippings overskuddsformål, både sett relativt til kommunens inntektsnivå og i absolute tall.

6.2 Hovedfunn, husholdninger

Hvilke sosioøkonomiske variabler kan forklare forskjeller i utgifter til pengespill hos norske konsumenter?

Med utgangspunkt i forbruksdata fra norske husholdninger, estimerer jeg en hindermodell som skiller mellom det binære valget spille / ikke spille; og kontinuerlige spilleutgifter gitt at husholdningen spiller. Cragg-modellen estimerer at inntekt, høyere utdanning og bosatt i Agder eller Rogaland innvirker negativt på sannsynligheten for at en husholdning deltar i pengespill fra Norsk Tipping. Jeg estimerer høyere sannsynlighet for spill med økt alder, og finner negativt estimat for alder kvadrert, altså er *veksten* avtakende med alder. Samtidig tolker jeg fra McFaddens pseudo R-kvadrat at modellens første trinn i liten grad kan predikere hvilke husholdninger som spiller. Modellens andre trinn, som estimerer utgifter til pengespill gitt at husholdningen spiller, har ingen signifikante parameterestimater. Når jeg kombinerer disse trinnene og beregner gjennomsnittlig partiellevirkning av inntekt på utgifter til pengespill, finner jeg at effekten ikke er signifikant forskjellig fra 0. I husholdningene ser det altså ut til at utgifter til pengespill i liten grad er korrelert med inntekt.

Flere fellesgoder finansieres av pengespill. Bryter dette med prinsipper om progressive eller proporsjonale skatter?

Når jeg kun ser på korrelasjonen mellom inntekt og utgifter til pengespill, uten å ta med øvrige forklaringsvariabler, finner jeg ikke et monotont voksende eller fallende mønster i pengespillutgifter for ulike inntektskvintiler i absolutte tall. Derimot finner jeg at husholdninger med lavere inntekter i større grad bidrar til Norsk Tippings overskuddsformål, relativt til eget inntektsnivå. Dette styrker hypotesen om pengespill som en regressiv form for skattlegging.

6.3 Sprik mellom kommune- og husholdningsdata

Regresjonsanalysen fra kommunene viser negativt parameterestimat for inntekt. Dette finner jeg ikke igjen i analysen av husholdninger. Her må det bemerkes at vi benytter to ulike inntektsvariabler, og det er derfor ikke uventet at vi får ulike estimater. Parameterestimatet for logaritmen av medianinntekt er ikke det samme som inntektselastisiteten; dette begrepet er knyttet til konsumentnivå og kan ikke uten videre finnes ved hjelp av aggregerte kommunedata. Samtidig er det ikke åpenbart hvordan vi skal forklare at vi har høyere spilleomsetning i lavinntektskommuner, dersom dette ikke skyldes inntektsnivået til enkelthusholdningene i kommunen. Oppgaven har diskutert svakheter med datasettet for husholdninger, og peker spesielt på problematikken rundt selvrapporing av utgifter til skambelagte aktiviteter som pengespill.

6.4 Videre forskning

Som diskutert i avsnitt 4.2 registreres alle utgifter til pengespill fra Norsk Tipping på et unikt spillernummer, som igjen kan knyttes til personnummer for alle spillere gjennom identifiseringstjenesten Buypass. I teorien kan forskeren da kunne få et datasett om spilleutgifter på konsumentnivå uten problemer med seleksjon og underrapportering. Spilledata kan kombineres med data som personinntekt, alder og bosted fra Skatteetatens databaser av en institusjon som kan behandle og anonymisere sensitive persondata. Korrelasjonen mellom personinntekt og utgifter til pengespill vil kunne undersøke spørsmål (2), altså hvorvidt den implisitte skatten som innkreves via pengespill rammer regressivt i befolkningen. Dersom det er mulig å inkludere øvrige sosioøkonomiske variabler som arbeidsstatus og utdanningsnivå fra andre offentlige registre, vil også spørsmål (1) kunne besvares med utgangspunkt i observert konsumentadferd.

7 Litteratur

- Beckert, J. og Lutter, M. (2009). The inequality of fair play: Lottery gambling and social stratification in Germany. *European Sociological Review*, **25**, 475 – 488.
- Bernoulli, D. (1954). Exposition of a new theory on the measurement of risk. *Econometrica*, **22**, 23 – 36.
- Burke, W. J. (2009). Fitting and interpreting Cragg's tobit alternative using Stata. *The Stata Journal*, **9**, 584 – 592.
- Clotfelter, C. T. og Cook, P. J. (1987). Implicit taxation in lottery finance. *National Tax Journal*, **40**, 533 – 546.
- Cragg, J. G. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica*, **39**, 829 – 844.
- The Economist: *Online Gambling. You bet.* (2010). Leder i trykket utgave, 8. juli 2010.
URL: <http://www.economist.com/node/16539402>. [Lesedato 1.5.2012]
- Fekjær, H. O. (2002). *Spillegalskap. Vår nye landeplage*. Oslo: Gyldendal.
- Friedman, M. og Savage, L. J. (1948). The utility analysis of choices involving risk. *Journal of Political Economy*, **56**, 279 – 304.
- Grønn, E. (2005). *Anvendt mikroøkonomi*. Oslo: Cappelen Akademisk Forlag.
- Haarr, T. K. og Krogstad, A. (2011). Myten om den norske kultureliten. *Sosiologisk Tidsskrift*, **19**, 6 – 28.
- Holme, A. (2004). *Matematikkens historie*. Bergen: Fagbokforlaget.
- Holmøy, A. (2008). *Forbruksundersøkelsen 2005. Dokumentasjonsrapport*. NOT 2008/9, Statistisk sentralbyrå.
- Jørsel, M. B. (2003). *Ludomani. Ikke flere innsatser, tak!* København: Gyldendal.

Kahneman, D og Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, **47**, 263 – 292.

Katz, V. J.(1998). *A history of mathematics – an introduction. Second Edition*. New York: Addison-Wesley.

Keen, M. (1986). Zero expenditures and the estimation of Engel curves. *Journal of Applied Econometrics*, **1**, 277 – 286.

Kwang, N. Y. (1965). Why do people buy lottery tickets? Choices involving risk and the indivisibility of expenditure. *Journal of Political Economy*, **73**, 530 – 535.

Lee, S. (1992). *Malcolm X*. New York: Warner Bros Pictures.

Lotteri- og stiftelsestilsynet. (2010). *Årsmelding 2010*.

URL: <https://lottstift.no/lotteritilsynet/files/2011/08/arsmelding-2010.pdf> [Lesedato 1.5.2012]

Lund, I. og Nordlund, S. (2003). *Pengespill og pengespillproblemer i Norge. SIRUS rapport 2/2003*. Oslo: Statens institutt for rusmiddelforskning.

Markowitz, H. (1952). The utility of wealth. *Journal of Political Economy*, **60**, 151 – 158.

Marshall, A. (2009). *Principles of economics: Unabridged Eighth edition*. New York: Cosimo.

Merverdiavgiftsloven. (2009). LOV 2009-06-19 nr 58: Lov om merverdiavgift.

URL: <http://www.lovdata.no/all/hl-20090619-058.html> [Lesedato 1.5.2012]

Norsk Tipping. (2006). *Årsrapport 2006*. Hamar: Norsk Tipping.

Norsk Tipping. (2009). *Års- og samfunnsrapport 2009*. Hamar: Norsk Tipping.

Norsk Tipping. (2010). *Års- og samfunnsrapport 2010*. Hamar: Norsk Tipping.

NOU (1997). *Norsk offentlig utredning 1997: 14. Spillet om pengene*. Oslo: Statens Forvaltningstjeneste.

- NOU (2003). *Norsk offentlig utredning 2003: 9. Skatteutvalget. Forslag til endringer i skattesystemet*. Oslo: Statens Forvaltningstjeneste.
- Ore, O. (1953). *Cardano. The gambling scholar*. Princeton, NJ: Princeton University Press.
- Pérez, L. og Humphreys, B. (2011). The income elasticity of lottery: New evidence from micro data. *Public Finance Review*, **39**, 551 – 570.
- Pran, K. R. og Ukkelberg, Å. (2010). *Spillevaner og spilleproblemer i befolkningen 2010*. Oslo: Synovate Norge.
- Schwartz, D. G. (2006): *Roll the bones. The history of gambling*. New York: Gotham Books.
- Silberberg, E. og Suen, W. (2001). *The structure of economics. A mathematical analysis. Third Edition*. New York: McGraw-Hill.
- Stewart, J. (2009). Tobit or not tobit? *U.S Bureau of Labor Statistics Working Paper 432*.
URL: <http://www.bls.gov/ore/pdf/ec090100.pdf> [Lesedato 1.5.2012]
- Suits, D. B. (1977). Measurement of tax progressivity. *The American Economic Review*, **67**, 747 – 752.
- TNS Gallup. (2012). Nyhetsbrev uke 16. URL: <http://www.tns-gallup.no/?did=9101560>
[Lesedato 1.5.2012]
- Varian, H. R. (1992). *Microeconomic Analysis. Third edition*. New York: Norton.
- Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Cambridge, Massachusetts: The MIT Press.
- Wooldridge, J. M. (2009). *Introductory econometrics. A modern approach. Fourth edition*. London: South-Western.
- Worthington, A. C. (2001). Implicit finance in gambling expenditures: Australian evidence on socioeconomic and demographic tax incidence. *Public Finance Review*, **29**, 326 – 342.

Øren, A. og Bakken, I. J. (2007). *Pengespill og pengespillproblem i Norge 2007*. Oslo: SINTEF Helse.

Øren, A. og Leistad, L. (2010). *Endringer i den norske befolkningens pengespillvaner og påvirkning av spilleautomatforbudet*. Trondheim: SINTEF Teknologi og samfunn.

Vedlegg 1

Snittomssetningen for lottospillere er 2179 kr per år (Norsk Tipping 2009), som tilsvarer 545 rekker a 4 kroner. I løpet av 60 år har en gjennomsnittsspiller da spilt 32 700 rekker.

Sannsynligheten for å få 7 rette i Lotto er gitt ved

$$P(7 \text{ rette}) = 1/\binom{34}{7} = 1/5379616$$

Sannsynligheten for at den typiske lottospilleren ikke får 7 rette i livet er gitt ved

$$P(\text{Ikke 7 rette}) = \left(\frac{5379615}{5379616}\right)^{32700} = 0,9939$$

Frem til 2012 var gevinstandelen til 7 rette 44,8 % av det totale gevinstbeløpet. (Norsk Tipping 2009). Dermed kan vi finne et uttrykk for forventningsverdien, gitt at spilleren ikke får 7 rette:

$$E(\text{Innsats} = I | \text{Ikke 7 rette}) = I * 0,5 * (1 - 0,448) = I * 0,276$$

Vi ser altså at 994 av 1000 normalspillere kun kan forvente å vinne tilbake 28 % av innsatsen sin. 6 av 1000 kan forvente å vinne betydelig mer. Videre kan vi se på sannsynligheten for å få nest høyeste gevinst:

$$P(6 + 1 \text{ rette}) = \binom{7}{6} * \binom{3}{1} / \binom{34}{7} = 21/5379616$$

Sannsynligheten for at den typiske lottospilleren verken får 7 eller 6+1 rette i livet er gitt ved

$$P(\text{Ikke 7 rette}) * P(\text{Ikke 6 + 1 rette}) = \left(\frac{5379615}{5379616}\right)^{32700} * \left(\frac{5379595}{5379616}\right)^{32700} = 0,8748$$

11,5 % av gevinstene går til 6+1 rette, som betyr at 875 av 1000 spillere vil ha forventingsverdi lik

$$E(I | \text{Ikke 7 rette, ikke 6 + 1 rette}) = I * 0,5 * (1 - 0,448 - 0,115) = I * 0,216$$

Vedlegg 2

Stata 12 er benyttet til alle estimeringer.

Do-fil, husholdninger.

```
use \\s-file\Home\st\terjko\Master_thesis\Data\Forbruksundersokelsen_2006\kritere2006.dta
keep LOEPENR G_INNT SFUTGIFT UTDANNING HTYPE_3 HPKJOENN HPALDER ANTYRK LANDSD
rename LOEPENR loepenr
merge loepenr using "\\s-file\Home\st\terjko\Master_thesis\Data\Forbruksundersokelsen_2006\f2006rec3ny.dta", unique sort
append using "\\s-file\Home\st\terjko\Master_thesis\Data\FU2005\FU_2005Full.dta"
keep G_INNT SFUTGIFT UTDANNING HTYPE_3 HPKJOENN HPALDER ANTYRK LANDSD vekt verva745 verva746 verva747
replace UTDANNING=99 if UTDANNING==.
gen grunnskole = 1 if UTDANNING <30
label variable grunnskole "=1 hvis grunnskole som høyeste fullførte utdanning"
replace grunnskole=0 if grunnskole==.
gen hoyereutd = 1 if UTDANNING >60 & UTDANNING<99
label variable hoyereutd "=1 hvis høyskole-/universitetsutdanning"
replace hoyereutd=0 if hoyereutd==.
gen ikkepar=1 if HTYPE_3==1 | HTYPE_3==5
label variable ikkepar "=1 hvis enslig eller aleneforelder"
replace ikkepar=0 if ikkepar==.
gen arbeidsledig = 1 if ANTYRK==0
label variable arbeidsledig "=1 hvis ingen arbeidsinntekt i husholdningen"
replace arbeidsledig=0 if arbeidsledig==.
gen agdrog=1 if LANDSD==4
replace agdrog=0 if agdrog==.
label variable agdrog "=1 hvis bosatt i Agder eller Rogaland"
gen aldersq=HPALDER*HPALDER
label variable aldersq "Hovedperson alder kvadrert"
sum verva745 G_INNT arbeidsledig grunnskole hoyereutd HPALDER agdrog

*finne antall med pengespillutgifter
gen spill = verva745+verva746+verva747
label variable spill "Samlet utgift til pengespill"
preserve
drop if spill==0
display (1-1292/2089)
restore

*konstruerer dummyvariabel =1 hvis utgifter til NT pengespill
gen w=1 if verva745>0
replace w=0 if w==.
label variable w "=1 hvis verva745>0"
drop if G_INNT==. /*får problemer med bootstrapping senere hvis vi har missing values*/
craggit w G_INNT arbeidsledig grunnskole hoyereutd ikkepar HPALDER aldersq agdrog, second(verva745 G_INNT
arbeidsledig grunnskole hoyereutd ikkepar HPALDER aldersq)

*sjekk hvis vi tar bort storspillere
preserve
drop if verva745>30000
craggit w G_INNT arbeidsledig grunnskole hoyereutd ikkepar HPALDER aldersq agdrog, second(verva745 G_INNT
arbeidsledig grunnskole hoyereutd ikkepar HPALDER aldersq)
restore

*tilbake til alle observasjoner
quietly craggit w G_INNT arbeidsledig grunnskole hoyereutd ikkepar HPALDER aldersq agdrog, second(verva745
G_INNT arbeidsledig grunnskole hoyereutd ikkepar HPALDER aldersq)

*beregne predicted verdier og partielleff for hver obs
predict xlg, eq(Tier1)
predict x2b, eq(Tier2)
predict sigma, eq(sigma)
gen Pw0=1-normal(xlg)/*sannsynlighet for y=0*/
gen Pw1=normal(xlg)
gen IMR=normalden(x2b/sigma)/normal(x2b/sigma) /*inverse mills ratio*/
gen Eyyx2=x2b+sigma*IMR /*forventningsverdi for verva745 gitt verva745>0*/
gen Eyxlx2=normal(xlg)*(x2b+sigma*IMR) /*ubetinget forventningsverdi for verva745*/

/*partielleffekt av inntekt på sannsynligheten for verva745>0*/
gen dPw1_dG_INNT=[Tier1]_b[G_INNT]*normalden(xlg)

/*partielleffekt av inntekt på verva745, gitt verva745>0*/
gen dEyyx2_dG_INNT=[Tier2]_b[G_INNT]*(1-IMR*(x2b/sigma+IMR))

/*partielleffekt av inntekt på ubetinget forventningsverdi av verva745*/
gen dEy_dG_INNT=[Tier1]_b[G_INNT]*normalden(xlg)*(x2b+sigma*IMR)+[Tier2]_b[G_INNT]*normal(xlg)*(1-
IMR*(x2b/sigma+IMR))

*kalkulere APE - gjennomsnitt partielleffekt av inntekt, mean = APE
```

```

sum dPw1_dG_INNT dEyyx2_dG_INNT dEy_dG_INNT

*finn andelen med høye/lave inntekter
preserve
drop if G_INNT<200000
display 195/2087
drop if G_INNT>800000
display 171/2087
restore

*finn pseudoR2 fra probit estimat
probit w G_INNT arbeidsledig grunnskole hoyereutd ikkepar HPALDER aldersq agdrog

*lage tabell, inntektskvintiler
preserve

sort G_INNT
gen i=_n
gen first = G_INNT if i < 418
gen second = G_INNT if i>417 & i<836
gen third = G_INNT if i>835 & i<1253
gen fourth = G_INNT if i>1252 & i<1671
gen fifth = G_INNT if i>1670

gen fi = 1 if first>-1000000 & first!=.
gen se = 1 if second>0 & second!=.
gen th = 1 if third>0 & third!=.
gen fo = 1 if fourth>0 & fourth!=.
gen ff = 1 if fifth>0 & fifth!=.

gen spillfirst = verva745*fi
gen spillsecond = verva745*se
gen spillthird= verva745*th
gen spillfourth= verva745*fo
gen spillfifth = verva745*ff

sum spillfirst spillsecond spillthird spillfourth spillfifth
sum first second third fourth fifth

*konfidensintervall for middelverdier i hver gruppe
mean spillfirst
mean spillsecond
mean spillthird
mean spillfourth
mean spillfifth

*er det signifikant forskjell mellom gruppene?
gen gruppe = 1 if fi==1
replace gruppe=2 if se==1
replace gruppe=3 if th==1
replace gruppe=4 if fo==1
replace gruppe=5 if ff==1
oneway verva745 gruppe, t bon

restore

*lage tabell over spillkvintiler
preserve

drop if verva745==0
sort verva745
gen i=_n
gen first = verva745 if i < 156
gen second = verva745 if i>155 & i<311
gen third = verva745 if i>310 & i<466
gen fourth = verva745 if i>465

gen fi = 1 if first>-1000000 & first!=.
gen sec = 1 if second>0 & second!=.
gen th = 1 if third>0 & third!=.
gen fo = 1 if fourth>0 & fourth!=.

gen spillfirst = verva745*fi
gen spillsecond = verva745*sec
gen spillthird= verva745*th
gen spillfourth= verva745*fo

sum spillfirst spillsecond spillthird spillfourth
restore

*standardfeil dPw1_dG_INNT med bootstrapping
program define APEboot1, rclass
preserve
craggit w G_INNT arbeidsledig grunnskole hoyereutd ikkepar HPALDER aldersq agdrog, second(verva745 G_INNT
arbeidsledig grunnskole hoyereutd ikkepar HPALDER aldersq)
predict bsx1g, eq(Tier1)
predict bsx2b, eq(Tier2)
predict bssigma, eq(sigma)

```

```

generate bsIMR=normalden(bsx2b/bssigma)/normal(bsx2b/bssigma)
generate bsdPw1_dG_INNT = [Tier1]_b[G_INNT]*normalden(bsx1g)
summarize bsdPw1_dG_INNT
return scalar ape1_G_INNT=r(mean)
matrix ape1_G_INNT=r(ape1_G_INNT)
restore
end
bootstrap ape1_G_INNT=r(ape1_G_INNT), reps(100): APEboot1

*standardfeil dEyyx2_dG_INNT med bootstrapping
program define APEboot2, rclass
preserve
craggit w G_INNT HPALDER aldersq ikkepar grunnskole hoyereutd arbeidsledig agdrog, second(verva745 G_INNT
HPALDER aldersq ikkepar grunnskole hoyereutd arbeidsledig)
predict bsx1g, eq(Tier1)
predict bsx2b, eq(Tier2)
predict bssigma, eq(sigma)
generate bsIMR=normalden(bsx2b/bssigma)/normal(bsx2b/bssigma)
generate bsdEyyx2_dG_INNT = [Tier2]_b[G_INNT]*(1-bsIMR*(bsx2b/bssigma+bsIMR))
summarize bsdEyyx2_dG_INNT
return scalar ape2_G_INNT=r(mean)
matrix ape2_G_INNT=r(ape2_G_INNT)
restore
end
bootstrap ape_G_INNT=r(ape2_G_INNT), reps(100): APEboot2

*standardfeil dEy_dG_INNT med bootstrapping:
program define APEboot, rclass
preserve
craggit w G_INNT arbeidsledig grunnskole hoyereutd ikkepar HPALDER aldersq agdrog, second(verva745 G_INNT
arbeidsledig grunnskole hoyereutd ikkepar HPALDER aldersq)
predict bsx1g, eq(Tier1)
predict bsx2b, eq(Tier2)
predict bssigma, eq(sigma)
generate bsIMR = normalden(bsx2b/bssigma)/normal(bsx2b/bssigma)
generate bsdEy_dG_INNT = [Tier1]_b[G_INNT] * normalden(bsx1g) * (bsx2b+bssigma*bsIMR) + [Tier2]_b[G_INNT] *
normal(bsx1g) * (1-bsIMR*(bsx2b/bssigma+bsIMR))
summarize bsdEy_dG_INNT
return scalar ape_G_INNT=r(mean)
matrix ape_G_INNT=r(ape_G_INNT)
restore
end
bootstrap ape_G_INNT=r(ape_G_INNT), reps(100): APEboot

```

Do-fil, kommuner

```

import excel "\\s-file\Home\ST\terjko\Master_thesis\Data\Data_kommuner\data_kommuner.xls", sheet("Datsett
masteroppgave") cellrange(A2:S431)
rename (A B C D E F G H I J K L M N O P Q R S) (kom oms inn fol kvi tet led gru uni kul sos pen ufø nor
al6t19 a20t44 a45t66 a67t79 a80pl)
label variable kom "Kommune"
label variable oms "Omsetning Norsk Tipping per innbygger"
label variable inn "Inntekt etter skatt, median"
label variable fol "Folkemengde 1. jan"
label variable kvi "Antall kvinner, 1. jan"
label variable tet "Andel bosatt i tettbebygd strøk"
label variable led "Andel registrert helt arbeidsledige, prosent"
label variable gru "Innbyggere med grunnskole som høyeste fullførte utdanning"
label variable uni "Innbyggere med universitets- høyskoleutdanning"
label variable kul "Netto driftsutgifter for kultursektoren per innbygger"
label variable sos "Sosialstønad per innbygger 16 år og over"
label variable pen "Antall alderspensjonister"
label variable ufø "Antall uførepensjonister"
label variable nor "Nordkommune. =1 hvis kommunen ligger i Nordland, Troms, Finnmark"
label variable al6t19 "Antall innbyggere, 16-19 år"
label variable a20t44 "Antall innbyggere, 20-44år"
label variable a45t66 "Antall innbyggere, 45-66 år"
label variable a67t79 "Antall innbyggere, 67-79 år"
label variable a80pl "Antall innbyggere, 80 år eller eldre"

gen tetpro = tet*100
label variable tetpro "Prosentandel bosatt i tettbebygd strøk"

gen men = fol-kvi
label variable men "Antall menn, 1. jan"
gen menpro = (men/fol)*100
label variable menpro "Prosentandel menn"

gen grupro = (gru/fol)*100
label variable grupro "Prosentandel med grunnskole som høyeste fullførte utdanning"

gen unipro = (uni/fol)*100
label variable unipro "Prosentandel med universitets- høyskoleutdanning"

```



```

gen apepro = (pen/fol)*100
label variable apepro "Prosentandel alderpensjonister i kommunen"

gen ufopro = (ufø/fol)*100
label variable ufopro "Prosentandel uførepensjonister i kommunen"

gen ung = a16t19 + a20t44
label variable ung "Antall innbyggere, 16-44 år"
gen ungpro = (ung/fol)*100
label variable ungpro "Prosentandel innbyggere fra 16 til 44 år"

*Genererer fylkesdummyer med utg punk i kommunenr

gen i=_n
gen trø=1 if i>293 & i<343
replace trø=0 if trø==.
label variable trø "Kommune i Trøndelag"

gen ves=1 if i>172 & i<294
replace ves=0 if ves==.
label variable ves "Kommune på Vestlandet"

gen sør=1 if i>142 & i<173
replace sør=0 if sør==.
label variable sør "Kommune på Sørlandet"

sum oms inn tetpro led kul sos menpro grupro unipro apepro ufopro ungpro

*regresjon med lineære variabler
reg oms inn led grupro unipro tetpro menpro kul sos apepro ufopro ungpro nor
estat ovttest
estat hettest

twoway (scatter oms inn) /*sjekker funksjonell form*/

gen lninn = log(inn)
gen lnoms = log(oms)
label variable lninn "Log av medianinntekt"
label variable lnoms "Log av omsetning NT"

*regresjon med log-variabler
reg lnoms lninn led grupro unipro tetpro menpro kul sos apepro ufopro ungpro nor
estat ovttest
estat hettest
vif

*regresjon med landsdel-variabler
reg lnoms lninn led grupro unipro tetpro menpro kul sos apepro ufopro ungpro nor trø ves sør
estat ovttest
estat hettest
vif

*Har forsøkt diverse variabelkombinasjoner. Dropper nor og tetpro, tar med sør
reg lnoms lninn led grupro unipro menpro kul sos apepro ufopro ungpro sør
estat ovttest
estat hettest
vif

*"Spuriøs" regresjon
reg lnoms lninn
twoway (scatter oms inn) (lowess oms inn)

*Dropper kommunen med lavest medianinntekt
preserve
drop if inn<300000
twoway (scatter oms inn) (lowess oms inn)
restore

```