96/3 | Notater | 1996

Statistics Norway

Statistisk sentralbyrå

Inger Marie Smestad

Valg under usikkerhet:
En analyse av eksperimentdata basert på kvalitative valghandlingsmodeller

Notater

Forskningsavdelingen / Seksjon for mikroøkonometri

Inger Marie Smestad

Valg under usikkerhet: En analyse av eksperimentdata basert på kvalitative valghandlingsmodeller

Sammendrag:

I denne oppgaven diskuterer jeg med utgangspunkt i diskret valghandlingsteori modeller for valg mellom alternativer der utfallet er usikkert, såkalte lotterier. Jeg tar utgangspunkt i modellen Becker et.al. fra 1963. Jeg tolker for ulike versjoner av modellen, i hvilken grad og hvordan aktøren benytter seg av tilgjengelig informasjon. Noen av modellene er det vanskelig å skille mellom på teoretisk basis. Jeg benytter data fra en rekke eksperimenter der personer velger mellom lotterier, til å estimere ulike modellversjoner. Ut fra resultatene diskutere jeg hvilken modell som best gjenskaper observert valgadferd.

FORORD

Jeg har i løpet av 9. semester ved Norges Teknisk Høyskole gjennomført en hovedoppgave for Statistisk Sentralbyrå. Oppgaven er knyttet til Dagsviks forskningsprosjekt om "Strukturelle modeller for diskrete/kontinuerlige valg (008)", og det er planen at mine resultater skal kunne utnyttes i en videreføring av prosjektet. I oppgaven tar jeg for meg modeller for kvalitativ valghandling med vekt på valg mellom lotterier. Det har vært skrevet og forsket svært lite på dette området. Hovedbidragene ble gitt på 1960-tallet. Blant andre har Becker et.al.(1963) lansert modeller som var generaliseringer av en modell for sikre utfall. Utover dette kunne de ikke rettferdiggjøre de nye modellene teoretisk. Området er komplisert og omfattende. Det er imidlertid mange som har hjulpet meg i løpet av de tre månedene jeg har holdt på med oppgaven. Jeg vil først og fremst rette en stor takk til min veileder John Kr. Dagsvik. Han har hele tiden vært tålmodig og tatt seg god tid til å hjelpe og rettlede meg. Ved hjelp av hans tålmodighet og kompetanse, håper jeg at denne oppgaven kan sees på som et anstendig bidrag. Videre vil jeg takke Bjørn Helge Vatne ved Statistisk Sentralbyrå og professor Stein W.Wallace for all hjelp både faglig og praktisk. Denne oppgaven markerer slutten på NTH-studiet og jeg vil til slutt takke for en fin og lærerik tid.

INNHOLDSFORTEGNELSE

TABELLISTE

FIGURLISTE

INNLEDNING

Tradisjonell teori for individuell valghandling tar for seg hvordan et individ vil fordele sin inntekt på en gitt mengde goder. I slike situasjoner er godene delbare og individet velger en kontinuerlig tilpasning. I denne oppgaven vil jeg imidlertid ta for meg valgmengder som består av goder som ikke er delbare. Typiske eksempler på slike goder er bolig, arbeid, utdannelse og transportmiddel. Et valg mellom goder av denne typen sies å være **kvalitativt eller kategorisk**. Individet står ovenfor en diskret og endelig valgmengde der eksakt et valgalternativ velges. Thurstone gjennomførte i 1927 en del eksperimenter der han blant annet observerte forsøkspersoners respons på fysiske stimuli som høye lyder og gullbarrer. Forsøkspersonene ble spurt om å rangere tonene etter styrke og gullbarrene etter tyngde. Thurstone observerte i stor grad inkonsistent rangering dersom tonene eller barrene ble for like. Dersom forsøkspersonene ble spurt om å velge en av gullbarrene, var det apriori rimelig å anta at de ville velge den de vurderte som tyngst, men basert på at de ikke er i stand til å evaluere hvert alternativ konsistent, ville et valg til en viss grad være tilfeldig. Det ville være tilfeldig i den forstand at forsøkspersonen ikke alltid ville velge den tyngste barren. En mulig tilnærming for å modellere et slikt inkonsistent valg, er å spesifisere opplevd tyngde av en gullbarre som en stokastisk størrelse. Thurstones eksperiment illustrerer en grunnleggende antagelse i forbindelse med stokastiske valghandlingsmodeller. Antagelsen går ut på at **verdien av et alternativ er en stokastisk størrelse**. Jeg har tatt utgangspunkt i bøker skrevet av blant andre Ben-Akiva og Lerman (1985), Train (1984) og Cramer (1991) når jeg i første del av oppgaven, presenterer teorien for denne typen modeller.

Det finnes mange eksempler på situasjoner der **utfallet av et valg er usikkert.** Et typisk eksempel på en slik situasjon er valg mellom å kjøpe et lodd i ulike lotterier. Modellen må i slike tilfeller kunne hanskes med usikkerhet i to sammenhenger. Det oppstår det som før usikkerhet i forbindelse med hvilket alternativ/lotteri individet vurderer som best. I tillegg er det usikkert hva individet vil sitte igjen med etter å ha valgt et lotteri. På bakgrunn av dette er det naturlig å dele kvalitative valghandlingsmodeller inn i to grupper. Jeg vil kalle dem for **kvalitative valghandlingsmodeller for valg med sikre utfall** og **kvalitative valghandlingsmodeller for lotterier**. I denne oppgaven bygger jeg opp et modellverk av den siste typen, det vil si for valgsituasjoner der utfallet av valget ikke er gitt i beslutningsøyeblikket. De største bidragene til et slikt modellverk ble gitt på 1960-tallet. Først ute var Becker et.al i 1963. Begge presenterte en modell for lotterier som inneholdt en modell for sikre utfall som et spesialtilfelle. Utover det ga de ingen teoretisk
begrunnelse. **I denne oppgaven vil jeg forsøke å gi en slik begrunnelse ved å ta utgangspunkt i modellen til Becker et.al. (1963b)**.

Oppgaven er lagt opp som følger. I **kapittel 1** forsøker jeg å motivere for kvalitative valghandlingsmodeller for valg med sikre utfall ved å innføre en stokastisk nytte. I **kapittel 2** går jeg konkret inn på to typer modeller, Logit og Probit, med hovedtyngden på de førstnevnte. Jeg tar for meg hvordan en modell bør spesifiseres og går inn på spesielle egenskaper ved den. Til nå har jeg kun referert eksisterende og kjent teori. I **kapittel 3** gjør jeg spranget over til kvaltitative valghandlingsmodeller for lotterier. Her ligger hovedtyngden av oppgaven. Kapittel 1 og 2 kan sees på som en nødvendig bakgrunn for å forstå det som nå kommer. I kapittel 3 tar jeg for meg hvordan og på grunnlag av hvilke data et individ vurderer et lotteri. Jeg kaller dette individets vurderingsmetode. Med utgangspunkt i Becker's modell (1963) og Dagsvik (1995) forsøker jeg å bygge opp en testbar modell som inneholder flere vurderingsmetoder. Notasjonen jeg bruker i oppgaven følger oppsummert i vedlegg A. I **kapittel 4** beskriver jeg estimeringsprosessen der sannsynlighetsmaksimering er et stikkord. Dette kapitlet er ment å gi leseren innsikt i statistiske verktøy og metoder som jeg senere bruker. Det inneholder ingen nye bidrag i forbindelse med modellteorien for lotterier. En stor del av oppgaven består i å formulere og begrunne en modell for lotterier både på et teoretisk og empirisk grunnlag. I **kapittel 5** beskriver jeg eksperimentet som danner grunnlaget for datamaterialet og resultatet av estimeringen. Spørreskjemaet og datamaterialet er å finne henholdsvis i vedlegg D og E.

Teoridelen, kapittel 1-3, er skrevet for å gi en innføring i kvalitative valghandlingsmodeller generelt og de mer ukjente modellene for lotterier spesielt. For å kunne forstå denne teorien forutsetter jeg at leseren har kjennskap til nytteteori og generell statistikk. Datamaterialet er i første rekke bearbeidet for forsker John K. Dagsvik ved Statistisk Sentralbyrå. I oppgaven bruker jeg det til å teste, eventuelt sammenligne de beskrevne modellene for lotterier.

1. TEORI FOR KVALITATIV VALGHANDLING

1.1 BEGREPER OG BAKGRUNN

I tradisjonell teori for individuell valghandling er godene delbare. Det vil si at konsumenten velger en kontinuerlig tilpasning. Sagt på en annen måte velger ikke individet ut ett av flere goder. Det velger mengden av hvert av dem. Det finnes imidlertid en del situasjoner der det er naturlig for konsumenten å velge kun et gode. Eksempler på dette er valg av livspartner, jobb, utdannelse og bolig. I slike valgsituasjoner kalles alternativene **kvalitative eller kategoriske**. Beslutningstakeren, individet som har myndighet til å gjøre valget, kalles **aktør**. Aktøren kan være både et enkelt individ og en gruppe. Eksempler på grupper er en familie, en hovedforsamling og Stortinget.

I kvalitative valgsituasjoner kan alternativene formuleres slik at de blir **gjensidig utelukkende**. Det vil si at aktøren velger kun ett av alternativene. Ta for eksempel valg av ektefelle. Aktøren kan i teorien selvfølgelig velge to, men det er lite gjennomførbart i praksis her i Norge og dessuten forbudt ved lov. Prinsippet må derfor sies å være oppfyllt. Et annet eksempel er valg av bosted. Aktøren kan velge å bo alene eller sammen med venner/familie i leilighet, hus, på hotell, i telt og så videre.
I noen tilfeller er ikke alternativene gjensidig utelukkende i utgangspunktet. De må derfor omformuleres for å oppfylle prinsippet. En ukependler må eksempelvis kunne velge et alternativ som betyr at han bor på hybel i ukedagene og hjemme i hus med sin familie i helgene. Videre er mengden av alternativer **fullstendig**. Listen over boalternativer eller potensielle ektefeller kan bli svært lang, og det er viktig å definere den slik at den omfatter alle mulige valg aktøren kan gjøre. Aktøren vil følgelig alltid velge et alternativ som er med i mengden. Til slutt består mengden av et **endelig antall** alternativer. Selv om listen over boalternativer kan bli lang er den ikke uendelig.

I valgteorien skiller man mellom rammebetingelser og individuelle preferanser. I kvalitativ valghandlingsteori er **rammebetingelsene representert ved mengden av tilgjengelige alternativer og variable som karakteriserer dem**. Vi kan igjen trekke en parallell til tradisjonell mikroteori. Der er rammebetingelser representert ved området under budsjettlinjen. Området under budsjettlinjen er bestemt av konsumentenes inntekt og prisene på godene. Rammebetingelsene kan følgelig variere fra aktør til aktør. Et eksempel på dette ved kvalitative valg er ved valg av bolig. Tilgjengelige boalternativer varierer med hvor aktøren søker bolig geografisk. Det samme gjelder en variabel som pris. Boliger er jevnt over dyrere i byer enn i spredtbebygde strøk.

Hvert alternativ er karakterisert ved et sett av variable kalt **attributter**. De danner et grunnlag for vurdering av verdien av hvert enkelt alternativ slik at aktøren er i stand til å skille mellom dem. Typiske attributter ved valg av boform er kostnad; representert ved leie eller kjøpspris og størrelse av boligen. Ved valg mellom hovedretter ved et restaurantbesøk kan attributtene blant annet være pris, mengde og type (kjøtt, fisk, vegetar).

Hver aktør er karakterisert ved et sett av variable kalt **aktørvariable**. Aktørvariablene skal beskrive variasjonen mellom aktørene. Sivilstatus og inntekt er eksempel på slike variable som kan antas å virke inn på valg av bolig.

Valgteorien antar at det eksisterer en sammenheng mellom rammebetingelser, aktørvariable og det valg aktøren gjør. For å finne denne sammenhengen baserer valgteorien seg på **nytteteorien**.

Et alternativ har en verdi for aktøren som er gitt ved en nytteindeks

$$U_{in} = f_n(w_n^*, s_i^*) \tag{1.1}$$

der w_n* er aktørvariable og s_i* er attributter. For hvert alternativ i som er valgbart for aktør n kan vi tenke oss en vektor med attributter. For hver aktør n finnes i tillegg en vektor med aktørvariable. Disse vektorene kombineres og reduseres til en skalar via nyttefunksjonen. Resultatet er N*S skalarer eller nytteverdier der S og N er henholdsvis antall alternativer og aktører. Nytten blir brukt til å rangere alternativene i forhold til hverandre. Nytteverdien er ordinal. Nærmere bestemt antas det at aktøren velger det alternativ som gir ham maksimal nytte. Han velger alternativ i dersom

$$U_{in} = \max_{j \in C_n}\left(U_{jn}\right) \tag{1.2}$$

der C_n er mengden av tilgjengelige alternativer for aktør n.

Ut fra en aktørs rammebetingelser og hans karakteristika er vi i teorien i stand til å predikere hans valg. Predikert og virkelig valg stemmer imidlertid ikke alltid i praksis. Årsaken til dette ligger blant annet i feilaktige antagelser om aktørens rammebetingelser og individuelle preferanser. Sagt på en annen måte er årsaken knyttet til feilaktige antagelser om valgmengden og nyttefunksjonen innbefattet dens parametre og variable. Jeg diskuterer årsaker til denne uoverensstemmelsen i neste avsnitt.

1.2 MOTIVASJON FOR STOKASTISK NYTTE

Nytteindeksen ble i avsnitt 1.1 beskrevet som

$$U_{in} = f_n(w_n^*, s_i^*)\ . \tag{1.3}$$

Nærmere spesifisert består vektorene

$$w_n^* = \left[w_n, w_n'\right]$$
$$s_i^* = \left[s_i, s_i'\right]$$

av observerbare , s_i og w_i, og uobserverbare , s_i' og w_i' , komponenter. Alle komponentene påvirker aktørens valg. Formålet med en modell er å forsøke å predikere dette valget. I modellen inngår imidlertid kun de observerbare komponentene. Modellen bygger derfor på et annet og mindre informasjons-grunnlag enn det aktøren tar hensyn til i sin valgprosess. Det resulterer i at aktørens valg kan være inkonsistent med det valg teorien predikerer. Denne inkonsistensen kan utdypes på to måter. Jeg vil i det følgende gå nærmere inn på dem begge.

1.2.1 Uobserverbare komponenter for forskeren

Det antas her at den sanne nytten knyttet til et alternativ i, U_{in}, er eksakt kjent for aktør n slik det fra hans synspunkt ikke er noen usikkerhet knyttet til verdien av alternativene. I aktørens nyttefunksjon inngår således både observerbare og uobserverbare komponenter. En forsker er interessert i å sette opp en modell for aktørens valg. Han må gjøre antagelser om aktørens nyttefunksjon siden han ikke kjenner dennes preferanser. Den antatte nyttefunksjonen kan avvike fra den sanne av flere årsaker.

Manski (1977) identifiserte fire:

- **Uobserverte attributter og aktørvariable.**

Aktøren tar hensyn til attributter og aktørvariable som ikke inngår i nyttefunksjonen spesifisert av forskeren. Sagt på en annen måte baserer aktøren sitt valg på et sett av variable w_n* og s_i* mens forskeren kun tar hensyn til w_n og s_i i modellen. Hvis vi betrakter det valget aktøren står ovenfor når han skal velge hovedrett på en restaurant, kan det for mange aktører være viktig å få poteter servert til hovedretten. Det faktum at alle retter ikke serveres med poteter, er derfor av avgjørende betydning. Dersom forskeren ikke har med dette attributtet i sin nyttefunksjon, vil han ikke fange opp all informasjon aktøren vektlegger i valgprosessen.

- **Uobserverte individuelle variasjoner.**

I praksis formuleres modeller som skal gjelde en gruppe aktører. Forskeren må følgelig ty til generaliseringer. En del av ulikheten mellom aktørene i en gruppe kan fanges opp gjennom aktør-variablene. I noen tilfeller kan imidlertid aktørene karakteriseres med lik verdi av aktørvariablene og allikevel ha ulike preferanser som forskeren ikke får fanget opp. Forskeren spesifiserer følgelig en nyttefunksjon f for en gruppe aktører som normalt ikke er eksakt lik den individuelle nyttefunksjonen f_n. I forbindelse med et restaurantbesøk kan aktørvariable være alder og inntekt. En aktør kan imidlertid ha en forkjærlighet for pasta mens en annen norsk husmannskost. Dette vil slå ut i ulik opplevd nytte for de to av for eksempel alternativet "Spaghetti Bolonaise" selv om de har samme alder og inntekt.

- **Målefeil og ufullstendig informasjon.**

Alle attributter eller aktørvariable er ikke like lette å observere og måle direkte. I noen tilfeller blir de indikert gjennom en annen variabel, som kan måles. De to nivåene er ikke alltid perfekt korrelerte, slik at aktørens nytteverdi vil avvike fra forskerens. Når det gjelder restauranteksempelet, trenger aktører ulik mengde av mat for å bli mette. En slik aktørvariabel, som jeg vil kalle magemål, virker inn på valg av matrett. Magemål kan ikke måles direkte, men det er rimelig å anta at det eksisterer en sammenheng mellom aktørens magemål og vekt. Da de to variablene ikke er perfekt korrelerte, kan resultatet bli inkonsistens mellom modellens predikerte og aktørens virkelige valg.

- **Feilspesifikasjon av nyttefunksjonens form.**

Nyttefunksjonen f er ukjent for forskeren. Det vil si han må anta en bestemt funksjonsform. Det er av bekvemlighetsårsaker vanlig å anta at nyttefunksjonen er lineær i de ukjente parametrene. I mange tilfeller er det imidlertid vanskelig å finne en teoretisk begrunnelse for en slik antagelse.

På bakgrunn av feilkildene beskrevet ovenfor, klarer forskeren i sin modell kun å fange opp en del av all den informasjon som inngår i aktørens valghandling. Nytten sies derfor å være **stokastisk for forskeren.**

1.2.2 Individuell inkonsistens

Thurstone gjennomførte i 1927 eksperimenter der han lot personer velge mellom gullbarrer med ulik tyngde. Samme person fikk gjentatte ganger velge i eksperimenter der det ble stadig vanskeligere å observere hvilken gullbarre som var tyngst. Thurstone mente at nytten til en gullbarre ifølge klassisk nytteteori burde være proporsjonal med pengeverdien. Det er i tillegg plausibelt og vanlig å anta at individet er nyttemaksimerende, noe Thurstone også gjorde. Ut fra disse to antagelsene forventet han ifølge teorien, at aktøren alltid ville velge den tyngste gullbarren. I sine forsøk fant imidlertid Thurstone at det ikke alltid stemte. Han observerte at

> **"there is some qualitative fluctuation from one occation to the next......for a given stimulus."**

(Thurstone 1927, s.280) og konkluderte derfor med at aktøren i mange situasjoner har problemer med å evaluere og rangere hvert enkelt alternativ konsekvent. Thurstone (1945) la frem en teori for dette fenomenet som i dag tolkes på følgende måte:

> **"the utilities are assumed to vary from moment to moment, and the decision process consists of the simple fixed rule of picking the alternative with largest *momentary* utility."**

(Edgell og Geisler 1980, s.266). Arbeidet til Thurstone ga han status som grunnleggeren av den stokastiske nytteteorien.

Jeg har nå motivert for stokastisk nytte på to måter. I kapittel 1.2.1 konkluderte jeg med at nytten var stokastisk for forskeren. I dette avsnittet sies nytten også å være **stokastisk for aktøren.** Denne forskjellen gir seg imidlertid ikke utslag i praksis. Uansett hvordan stokastisk nytte er motivert, er resultatet, modellen, den samme.

På bakgrunn av diskusjonen ovenfor er en deterministisk modell ofte uegnet til å predikere en aktørs valg. Thurstone var en av pionerene i arbeidet med å generalisere modellformuleringen slik at den tok hensyn til inkonsistens i aktørenes adferd. Løsningen han presenterte var å modellere nytten som en stokastisk, normalfordelt variabel. Den resulterende modellen blir idag kalt **Probit**modellen. Senere har det blitt introdusert alternative modeller. Den mest brukte er **Logit**modellen som jeg vil komme mer inn på senere.

2. KVALITATIVE VALGHANDLINGSMODELLER FOR VALG MED SIKRE UTFALL

Jeg vil i første omgang ta for meg situasjoner der aktøren er sikker på utfallet av valget i beslutningsøyeblikket. I tillegg vil jeg i første omgang forenkle situasjonen ytterligere ved å kun se på situasjoner der aktøren står ovenfor to mulige valg; en binær modell. Grunnleggende prinsipper blir på den måten enklere å forklare. Utvidelse til en multinomisk modell bygger på de samme prinsippene som for den binomiske modellen.

Kvalitativ valghandlingteori forutsetter som nevnt at det er en sammenheng mellom attributter, aktørvariable og aktørens valg. Sammenhengen mellom aktørens valg og de observerbare komponentene spesifiseres i den deterministiske delen av nyttefunksjonen. I avsnitt 2.1.1 går jeg nærmere inn på hvordan preferanser, attributter og aktørvariable kan spesifiseres i en modell. Videre må forskeren gjøre antagelser vedrørende den stokastiske delen eller restleddet. Han må ta stilling til hvilken fordeling av restleddet som på best mulig måte kompenserer for den ufullstendige informasjonen forskeren sitter inne med. Dette går jeg inn på i avsnitt 2.1.2.

For å illustre prinsippene modellen bygger på, vil jeg bruke et gjennomgående eksempel: Aktøren skal bestille hovedrett på en restaurant der menyen består kun av to retter; kokt laks og hamburger. Ingen andre retter er tilgjengelig. Rammebetingelsene er dermed satt.

Fra avsnitt 1.2 har vi at nytten består av en deterministisk og en stokastisk komponent. Dersom vi antar at disse to **delene er separable**, kan nytten skrives som

$$\tilde{U}_{in} = \tilde{u}_{in}(s_i, w_n)\tilde{\varepsilon}_{in}(s_i', w_n') = \tilde{u}_{in}\tilde{\varepsilon}_{in} \tag{2.1}$$

der $\tilde{u}_{in}$ er en deterministisk funksjon som antas kjent for forskeren på et sett av ukjente parametre nær og $\tilde{\varepsilon}_{in}$ en stokastisk variabel. Den stokastiske variabelen er et forsøk på å modellere tilsynelatende inkonsistente valg beskrevet i avsnitt 1.2. Nytteindeksen er som nevnt tidligere et ordinalt mål. Den kan kun brukes til å rangere alternativene. **Enhver monoton transformasjon av nytten er derfor tillatt siden den lar rangeringen være uforandret**. Vi skal dra nytte av denne egneskapen i flere sammenhenger, og det er derfor viktig å merke seg den. Valget forblir det samme dersom aktørens opplevde nytte beskrives av

$$U_{in} = \ln(\tilde{U}_{in}) = \ln(\tilde{u}_{in}) + \ln(\tilde{\varepsilon}_{in}) = u_{in} + \varepsilon_{in} \ . \tag{2.2}$$

Siden den modellerte nytten er stokastisk, gir ikke modellen en deterministisk prediksjon for hvilket alternativ aktøren vil velge. **Den kan høyst predikere sannsynligheten for at aktør n velger alternativ i fra mengden C_n**. Denne sannsynligheten er formelt definert ved

$$P_n(i;C_n) = P\left(U_{in} = \max_{j\in C_n}(U_{jn}) \right) . \tag{2.3}$$

Sagt med ord er sannsynligheten for at aktøren velger alternativ i, den samme som sannsynligheten for at U_{in} er den høyeste nytten knyttet til alternativene i C_n. For å kunne spesifisere denne sannsynligheten ytterligere er vi nødt til å gjøre videre antagelser om den deterministiske og den stokastiske delen av nyttefunksjonen, nemlig u_{in} og ε_{in}.

2.1 BINÆRE MODELLER

I en binær valgsituasjon kan aktøren velge mellom to alternativer.

2.1.1 Spesifikasjon av deterministisk komponent

Nytteverdiens deterministiske komponent er, som nevnt i avsnitt 1.1, en funksjon av attributter ved et alternativ og karakteristika ved aktøren. Disse er spesifisert av forskeren og samlet i hva jeg i det følgende kaller for en variabelvektor $\mathbf{x_{in}}$. Det er ofte antatt at nyttefunksjonen er linear i sine parametre $\boldsymbol{\beta}_\mathbf{n}$ slik at den er gitt ved

$$u_{in} = \beta_{1n}x_{in1} + \beta_{2n}x_{in2} + \dots + \beta_{Mn}x_{inM} = \boldsymbol{\beta}_\mathbf{n}^T\mathbf{x_{in}} \tag{2.4}$$

hvor M er antall variable. Som nevnt i avsnitt 1.2.1 finnes det ingen teoretisk begrunnelse for en bestemt spesifikasjon av nyttefunksjonen. Antagelsen om linearitet er gjort for at det skal være lett å estimere de ukjente parameterene. Ligning (2.4) bestemmes følgelig av to deler:

♦ **Variabelvektor $\mathbf{x_{in}}$.**

Variabelvektoren består av komponenter som varierer med alternativene, $\mathbf{s_i}$, og komponenter som varierer med aktørene, $\mathbf{w_n}$, det vil si henholdsvis **attributter** og **aktørvariable**. Videre er det ofte fornuftig å modellere såkalte **samspillsvariable** som varierer både med aktør og attributt. De kan modelleres som en kombinasjon mellom attributter og aktørvariable. Variabelvektoren kan følgelig skrives som

$$\mathbf{x_{in}} = \begin{bmatrix}\mathbf{s_i} & \mathbf{w_n} & \mathbf{h_{in}}(\mathbf{s_i}, \mathbf{w_n})\end{bmatrix}^T$$

der vektorfunksjonen $\mathbf{h_{in}}(.)$ beskriver en kombinasjon av $\mathbf{s_i}$ og $\mathbf{w_n}$. Forskeren søker en modell med aktørvariable som fanger opp mest mulig individuell variasjon. Samtidig skal attributtene uttrykke forskjellen mellom alternativene.

Antagelsen om lineær nyttefunksjon betyr **ikke** at den behøver å være lineær i sine variabler eller attributter. Vi kan godt tenke oss et attributt eller en aktørvariabel som er en transformasjon av en annen. Et vanlig eksempel i den forbindelse er å bruke

$$x_{inm}^* = \ln(x_{inm})$$

i modellen istedet for x_{inm} dersom det dreier seg om pengeverdier. Transformasjonen er motivert av at det i mange situasjoner er rimelig å anta at en økning på 1 krone i for eksempel lønn gir en større økning av nytte ved lave inntekter enn ved høye. Sagt mer presist observeres en avtagende marginal nytte. Omvendt gir

$$x_{inm}^* = x_{inm}^2$$

en økende marginal nytte for $x_{inj} > 0$.

Dersom jeg går tilbake til eksemplet om valg av hovedrett på en restaurant, er blant andre pris, mengde og type, attributter som kan regnes å ha innvirkning. Attributtet type må spesifiseres nærmere. En mulig inndeling er kjøtt, fisk og vegetarmat. Variabelen blir diskret og må tilordnes valgte verdier. Verdien av disse varierer kun med alternativene. Inntekt og alder er eksempel på aktørvariable som kan tenkes å ha betydning. For blant andre inntekt og pris kan det lønne seg å teste om en ln-transformasjon kan gi en bedre modell, det vil si bedre tilpasning til datamaterialet. Årsaken til at en slik transformasjon kan gi bedre tilpasning er som nevnt at det i sammenheng med variable målt i pengeverdi ofte observeres avtagende marginal nytte. Det vil si at en gitt inntekts-eller prisforskjell er mer utslagsgivende ved et lavt nivå enn et høyt. Det er også rimelig å anta en samspilleffekt mellom disse variablene. Det vil si at en aktør med høy inntekt legger mindre vekt på pris og omvendt. Samspillvariabelen

$$x_{inm} = \frac{\text{Pris for matrett i}}{\text{Inntekt til aktør n}}$$

kan derfor være en bedre forklaringsvariabel enn attributtet pris og aktørvariabelen inntekt hver for seg.

- **parametervektor $\boldsymbol{\beta}_n$.**

Parametervektoren er et uttrykk for **aktørens preferanser**. Sagt på en annen måte består den av et sett med vekter som er knyttet til de ulike variablene. Ideelt kan parametervektoren variere både med alternativ og aktør. En aktør kan sterkt vektlegge at matretten skal være billig mens det for en annen er typen som er viktigst. I praksis er det imidlertid ofte ikke mulig å bestemme en parametervektor for hver enkelt aktør fordi man sjelden har mer enn noen få observasjoner for hvert individ. Forskeren antar i den forbindelse at parametervektorene er like for alle aktører; $\boldsymbol{\beta} = \boldsymbol{\beta}_n$. Han antar med dette at alle aktører har de samme preferansene etter å ha kontrollert for aktørvariablene $\mathbf{w}_n$. Sagt på en annen måte influeres aktørene like mye av de samme variable. Antagelsen holder kun ved en homogen populasjon. Dersom gruppen som undersøkes er svært heterogen og variasjonen ikke kan uttrykkes gjennom aktørvariablene, bør gruppen deles i to eller flere undergrupper. Oppdelingen foregår ved hjelp av et segmenterings-kriterie. Alder er et eksempel på et slikt kriterie i forbindelse med valg av hovedrett på en restaurant. I hver aldersgruppe estimeres så en felles parametervektor. Denne prosessen kalles markedssegmentering. Ved markedssegmentering i sin mest ekstreme form defineres en parametervektor til hver enkelt aktør. Ved markedssegmentering spesifiseres ikke samspillet $\mathbf{h}_{in}(\mathbf{s}_i,w_t)$ mellom segmenteringsvariabelen w_t og de øvrige attributtene $\mathbf{s}_i$. Det er imidlertid mulig å spesifisere et slikt samspill. Parametervektoren $\boldsymbol{\beta}_n$ kan eksempelvis være lineært avhengig av segmenteringsvariabelen. Det vil si at

$$\mathbf{h}_{in}(\mathbf{s}_i,w_t) = (\boldsymbol{\beta}_{0n}+\boldsymbol{\beta}_{1n}w_{tn})^T\mathbf{s}_i$$

En antagelse om like parametervektorer er imidlertid svært vanlig.

2.1.2 Spesifikasjon av restleddsfordelingen

Vi må, i tillegg til å beskrive den deteministiske delen av nyttefunksjonen, også gjøre en antagelse om den stokastiske delen for å kunne estimere en modell. Ulik spesifisering av fordelingen til ε_i resulterer i forskjellige modeller. Restleddet representerer som nevnt i avsnitt 1.2.1 avviket mellom aktørens og forskerens nyttefunksjon. Alternativt kan det som nevnt i avsnitt 1.2.2 være et uttrykk for inkonsistens eller ufullstendig informasjon fra aktørens side.

Ved et valg mellom to alternativer, i og j , gjelder for begge tilfellene at:

$$P_n(i;\{i,j\}) = P(U_{in} > U_{jn}) = P(\varepsilon_{jn} - \varepsilon_{in} \leq u_{in} - u_{jn}) \tag{2.5}$$

noe som viser at fordelingsfunksjonen, $P_n(i;\{i,j\})$, kun er avhengig fordelingen til **differansen** mellom restleddene. For å kunne beregne $P_n(i;\{i,j\})$ må forskeren derfor anta en hensiktsmessig og plausibel fordeling av ε_{in} . Avhengig av denne antagelsen får han ulike modeller. Jeg vil i de følgende to avsnittene gå inn på de mest vanlige restleddsfordelingene; normal-og ekstremfordelingen.

2.1.2.1 Probitmodellen

I Probitmodeller er restleddene antatt å være **normalfordelte.** Restleddet antas å ha forventning null og en fordeling som er uavhengig av dendeterministiske delen u_{in}. Normalfordelingen er vist i Figur 2-1. Restleddene kan være korrelert med hverandre, noe som også ofte er tilfelle i praksis. Dette virker rimelig dersom vi ser tilbake på argumentasjonen for stokastisk nytte. De uobserverbare komponentene virker ofte inn på flere alternativer. Restleddene kan sees på som regressorer som er avhengig av de samme variable, om enn i ulik grad. Av dette følger en samvariasjon mellom regressorene. Antagelsen om normalfordelte restledd med forventning null resulterer i modellen

$$P_n(i;\{i,j\}) = P(U_{in} > U_{jn}) = P(\varepsilon_{jn} - \varepsilon_{in} \leq u_{in} - u_{jn}) = \Phi\left(\frac{u_{in} - u_{jn}}{\sigma}\right) \tag{2.6}$$

hvor

$$\sigma^2 = \mathrm{var}(\varepsilon_{jn} - \varepsilon_{in}) = \mathrm{var}(\varepsilon_{jn}) + \mathrm{var}(\varepsilon_{jn}) - 2\,\mathrm{cov}(\varepsilon_{jn}, \varepsilon_{in}) \tag{2.7}$$

og Φ er den standardiserte, kumulative normalfordelingen.

2.1.2.2 Logitmodellen

I Logitmodeller antas restleddene å være **identisk og uavhengig ekstremfordelt** av type I; såkalt Gumbelfordelt (vedlegg B). Dersom ε_{in} og ε_{jn} er Gumbelfordelt med samme middelverdi og varians, er ε_{jn}-ε_{in} **logistisk** fordelt med

$$P(\varepsilon_{jn} - \varepsilon_{in} < y) = F(y) = \frac{1}{1 + e^{-\mu y}} \tag{2.8}$$

der μ er en positiv skaleringsparameter som har tolkning som

$$\mathrm{var}(\varepsilon_{jn} - \varepsilon_{in}) = \frac{\pi^2}{3\mu^2}\,. \tag{2.9}$$

Modellantagelsene resulterer i valgsannsynligheten

$$P_n(j;\{i,j\}) = P(U_{in} < U_{jn}) = P(\varepsilon_{in} - \varepsilon_{jn} < -(u_{in} - u_{jn})) = \frac{1}{1 + e^{\mu(u_{in} - u_{nj})}} \tag{2.10}$$

eller

$$P_n(j;\{i,j\}) = \frac{e^{\mu u_{jn}}}{e^{\mu u_{jn}} + e^{\mu u_{in}}}\,. \tag{2.11}$$

Størrelsen av μ kan ikke bestemmes entydig i modellen. Det er derfor vanlig å anta at $\mu=1$. Antagelsen medfører ikke tap av generalitet. I stedet for å variere μ kan forskeren tilpasse parametrene i nyttefunksjonen. Dersom estimeringen resulterer i enten svært store eller svært små parametre i forhold til uforklart variasjon i datamaterialet, fåes to grensetilfeller:

$$\frac{\operatorname{var}(\varepsilon_{in}-\varepsilon_{jn})}{u_{in}}=\frac{\pi^2}{3u_{in}}\rightarrow\infty\Rightarrow P_n(i;\{i,j\})=0.5 \qquad \text{(kun tilfeldig valg)}$$

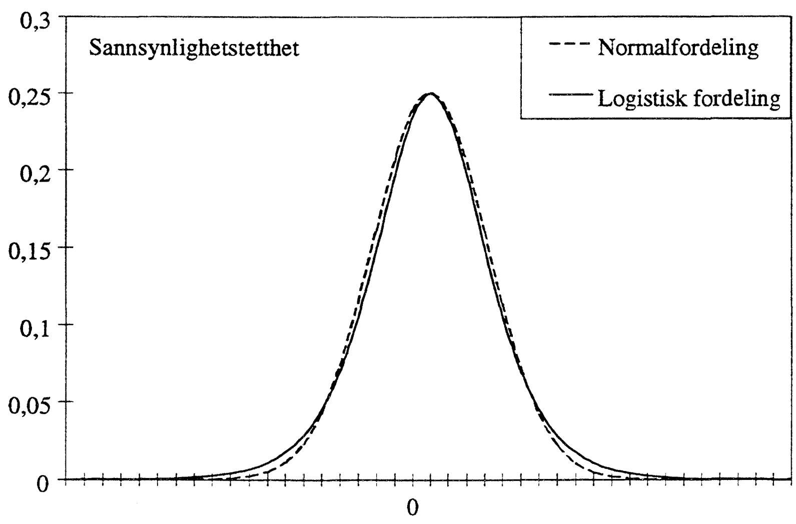
$$\frac{\operatorname{var}(\varepsilon_{in}-\varepsilon_{jn})}{u_{in}}=\frac{\pi^2}{3u_{in}}\rightarrow 0\Rightarrow P_n(i;\{i,j\})=\begin{cases}1 \text{ når } U_{in}>U_{jn}\\ 0 \text{ ellers}\end{cases} \qquad \text{(deterministisk valg).}$$

Følgelig bestemmer nivået på de deterministiske komponentene stigningen av fordelingsfunksjonen. Intuitivt stemmer dette. Jo større del av nytten forskeren mener han klarer å måle gjennom den deterministiske komponenten han antar, dess mer deterministisk blir modellen.

2.1.2.3 Valg av restleddsfordeling

Logit-og Probitmodellen er de to mest vanlige i forbindelse med binære kvalitative valghandlingsmodeller. Den logistiske fordelingen er av form svært lik normalfordelingen slik at modellene blir svært like. Begge fordelingene er vist i Figur 2-1. Forskjellen mellom de to fordelingene er størst i halene. Valgsannsynlighetene de to modellene predikerer, stemmer med andre ord bedre overens dersom de deterministiske komponentene ikke antar ekstreme verdier. I det binære tilfellet gir dermed modellene en tilnærmet lik prediksjon, og valg av restleddsfordeling er av mindre betydning.

Figur 2-1: Sammenligning av normal-og logistisk fordeling



I det multinomiske tilfellet finnes det argumenter både for å anta ekstremfordelte og normalfordelte restledd. I spesifisering av en modell er det viktig å reflektere over om de krav antagelsene innebærer, virkelig er oppfyllt. Logitmodellen innebærer uavhengig og identisk ekstremfordelte restledd. Uavhengighet behøver ikke alltid å være tilfelle, noe vi skal komme tilbake til i avsnitt 2.2. Dersom restleddene ikke kan sies å være uavhengige, er det derfor ofte vanlig å velge en Probitmodell. I situasjoner med kun to alternativer har denne modellen vært mye brukt. Situasjoner med flere enn to valgmuligheter fører imidlertid til modeller med svært komplisert regning siden valgsannsynlighetene ikke kan skrives på lukket form. Da er logitmodellen er mye enklere i slike tilfeller. Det finnes i slike tilfeller generaliserte Logitmodeller som løser på kravet om uavhengige restledd. "Nested Logit model" er et eksempel på en slik modell.

2.2 MULTINOMISK LOGITMODELL

Til nå har jeg, som illustrasjon, konsentrert meg om en modell med bare to valgmuligheter. Med Logitmodellen er det som nevnt, matematisk ikke noe i veien for utvidelser. Dersom C_n er en mengde bestående av S valgalternativer, er valgsannsynlighetene i en multinomisk Logitmodell gitt ved:

$$P_n(i;C_n) = \frac{e^{u_{in}}}{\sum_{t=1}^{S} e^{u_{tn}}} \quad . \tag{2.12}$$

Prinsippene for spesifikasjon av deterministisk og stokastisk del av nyttefunksjonen gjelder fremdeles[1].

2.2.1 Elastisiteter

Direkte- og krysselastisiteter[2] er interessante størrelser som sier noe om sammenhengen mellom sannsynlighetsfordelingen $P_n(i;C_n)$ og variablene i modellen. Dersom modellen skrives som

$$U_{in} = \beta^T \mathbf{x}_{in} + \varepsilon_{in} = \beta_1 x_{in1} + \ldots + \beta_M x_{inM} + \varepsilon_{in} \tag{2.13}$$

er direkte elastisitet definert som

$$E_{x_{ink}}^{P_n(i;C_n)} = \frac{\partial P_n(i;C_n)}{\partial x_{ink}} \frac{x_{ink}}{P_n(i;C_n)} = \frac{\partial \ln(P_n(i;C_n))}{\partial \ln(x_{ink})} = [1 - P_n(i;C_n)] x_{ink} \beta_k \tag{2.14}$$

og krysselastisitet som

$$E_{x_{jnk}}^{P_n(i;C_n)} = \frac{\partial P_n(i;C_n)}{\partial x_{jnk}} \frac{x_{jnk}}{P_n(i;C_n)} = \frac{\partial \ln(P_n(i;C_n))}{\partial \ln(x_{jnk})} = -P_n(j;C_n) x_{jnk} \beta_k \quad . \tag{2.15}$$

Et eksempel på en direkte elastisitet kan være relativ endring av sannsynligheten for at individ n skal velge laks når prisen for laks endres med 1 krone. Relativ endring av den samme sannsynligheten når prisen for en hamburger endres med 1 krone, er et eksempel på en krysselastisitet. Med en relativ endring mener jeg total endring i forhold til opprinnelig verdi.

Vi ser av (2.15) at effekten av en endring i en variabel x_{jnk} fordeler seg likt over alle alternativer. Det er et annet uttrykk for IIA-egenskapen som vi kommer inn på i neste avsnitt.

[1] For utledning av sannsynlighetfordelingen for multinomisk modell viser jeg til vedlegg C.
[2] Med elastisiteter menes her punktelastisiteter; de deriverte er evaluert i startpunkt.

2.2.2 Uavhengighet av irrelevante alternativer

Stengt tatt er Logitmodellen den eneste som oppfyller prinsippet om **uavhengighet av irrelevante alternativer,** oftest omtalt som IIA-prinsippet (Luce, 1959). Alle modeller som er basert på en antagelse om **uavhengig fordelte restledd**, har imidlertid **tilnærmet** IIA-egenskapen (Ben-Akiva og Lerman 1991, s 109).

Prinsippet innebærer at dersom vi isolert betrakter to alternativer, i og j, i en modell, skal forholdet mellom de respektive valgsannsynlighetene

$$\frac{P_n(i;C_n)}{P_n(j;C_n)} = r(u_{in}, u_{jn}) \tag{2.16}$$

være uavhengig av C_n. Altså skal **funksjonsverdien $r(u_{in},u_{jn})$ forbli uendret dersom rammebetingelser knyttet til andre alternativ enn i og j endres.** Et valg sies på bakgrunn av dette å være uavhengig av irrelevante alternativer.

For å illustrere hva prinsippet innebærer, la oss tenke oss en meny bestående av to alternativer; laks og hamburger. Jeg vil kalle dem henholdsvis alternativ 1 og 2. La oss betegne forholdet mellom de to valgsannsynlighetene med P(1)/P(2). Restauranten utvider så menyen med ett tredje alternativ; skinkesalat. **Dersom aktør n får velge fra den nye menyen, skal dette forholdet fortsatt være det samme og uavhengig av u_{3n}.**

Jeg nevnte innledningsvis at mengden av alternativer skulle være fullstendig. En fordel knyttet til alternativer med denne egenskapen er at **forskeren kan estimere modeller bygget på kun deler av den totale valgmengden.** I en del tilfeller kan forskeren være interessert i valgsannsynlighetene for kun en delmengde av alternativene. Dersom en finansinstitusjon er interessert i å undersøke sannsynligheten for valg av deres aksjefond i forhold til konkurrentenes, behøver de ikke ta alle andre spareformer med i betraktningen. Observasjoner av aktører som har valgt utenfor delmengden som undersøkes, utelates.

Prinsippet om IIA setter visse krav til alternativene i en valgsituasjon. Det finnes tilfeller der det ikke er rimelig å anta at prinsippet stemmer. Dersom menyen på vår restaurant utvides med ørret i stedet for skinkesalat, gjelder sannsynligvis ikke prinsippet om IIA. Årsaken er som følger. Alle modeller som er basert på en antagelse om **uavhengig fordelte restledd**, har tilnærmet IIA-egenskapen (Ben-Akiva og Lerman 1991, s 109). Som en følge av antagelsen om uavhengige restledd, er de sanne nytteverdiene knyttet til hvert alternativ også uavhengige. Sannsynligheten for at en nytteverdi er størst i en delmengde er dermed uavhengig av nytten til alternativene utenfor mengden. Det er rimelig å anta at sannsynligheten for å velge laks fra en delmengde som består av laks og hamburger, synker dersom ørret blir introdusert som et tredje alternativ. Alternativene laks og ørret har svært mange fellestrekk; fisk med tilnærmet samme smak og farge. Det rimelig å anta at de respektive nytteverdiene påvirkes av de samme uobserverte faktorer. Restleddene kan dermed ikke sies å være uavhengig fordelte. En modell med IIA-egenskapen egner seg følgelig ikke i denne situasjonen.

3. KVALITATIVE VALGHANDLINGSMODELLER FOR LOTTERIER

Til nå har jeg sett på modeller for valg med sikre utfall. Har aktøren bestemt seg for å bestille laks, har han også fått det. Det er imidlertid mulig å tenke seg en ny type valgalternativer hvis **utfall ikke er kjent for aktøren i beslutningstidspunktet**. Et typisk eksempel på en slik situasjon er når aktøren skal velge mellom ulike lotterier. Modeller som beskriver slike valgsituasjoner kalles følgelig for kvalitative valghandlingsmodeller for lotterier. Det oppstår usikkerhet i to trinn; i første omgang i forbindelse med valg av lotteri og deretter i forbindelse med utfallet av det valgte lotteriet.

Jeg vil illustrere den nye situasjonen med et eksempel. Samtidig vil jeg beskrive notasjonen. Utenfor butikken du er inne og handler i står det to loddselgere. De representerer to lotterier. Du har bestemt deg for å kjøpe et lodd. Mengden av mulige valg C består av to lotterier. Hvert lotteri har en mengde gevinster, D_1 og D_2. I det første lotteriet kan du vinne en fruktkurv eller trøstepremier og i det andre en reise til Gambia eller en weekend på Lillehammer Hotell. Av loddselgerne får du vite hvor mange lodd som selges i hvert lotteri. Ut fra dette er du i stand til å regne ut sannsynligheten for å få gevinst J i lotteri i. J er følgelig en stokastisk variabel. For å summere opp notasjonen, består mengden av mulige valg C av S lotterier. For lotteri $i \in C$ finnes det en mengde D_i bestående av K_i mulige utfall. Til hvert enkelt utfall og lotteri er det knyttet en nytte U_{iJ} der $J \in D_i$. Denne nytteverdien er tilsvarende som ved modeller for sikre utfall, avhengig av attributter ved utfallet, aktørvariable samt en stokastisk variabel. For å kunne evaluere et lotteri, må aktøren i tillegg til utfallsnyttene vite sannsynligheten g_{iJ} for et enkelt utfall J av et lotteri i. Vektoren

$$\mathbf{g}_i = \left[g_{i1} \;\; g_{i2} \cdots\cdots g_{iK_i}\right]^T \qquad i = 1,2,\ldots,S$$

antas kjent og lik for både aktør og forsker (modell). All denne informasjonen er allikevel ikke nok til at aktøren klarer å skille mellom lotteriene. Vi kan trekke en parallell til kvalitative valghandlingsmodeller for sikre utfall. Der fulgte aktøren en beslutningsregel om å velge det alternativ som ga maksimal nytte for å skille alternativene fra hverandre. Dersom, i tillegg til selve valget, også utfallet av valget er stokastisk, sier vi at aktøren tilsvarende må bruke en **vurderingsmetode** for å skille mellom lotteriene. Vurderingsmetoden beskriver måten aktøren kombinerer sine preferanser med den informasjonen som ligger i utfallssannsynlighetene. Jeg vil gå nærmere inn på dette i kapittel 3.2.

Dersom man sammenligner innsatsen satt inn på modeller for lotterier i forhold til for modeller for valg med sikre utfall, er den forsvinnende liten. Årsakene til dette er ikke mangel på bruksområder. Eksempel på et anvendelsesområde kan være valg mellom ulike sparingsformer. En aktør kan velge mellom et sett av sparingsalternativer; eksempelvis oppbevare sine penger i madrassen eller i form av bankinnskudd, aksjer, fonder eller forsikring. Hvert enkelt sparingsalternativ kan sees på som ett lotteri. Dersom vi betrakter en sparingshorisont på ett år, er det usikkert hvilken avkastning et sparingsalternativ vil gi. Vi kan tenke oss en kontinuerlig utfallssannsynlighetstetthet og en utfallsnytte knyttet til pengeverdien ett år etter sparestart. Med en slik modell vil det være mulig å predikere etterspørselen etter de forskjellige spareformene. Et annet mulig anvendelsesområde er for å predikere etterspørselen etter en bestemt utdannelse. For en gitt utdannelse finnes det en gitt mengde av jobbmuligheter. Sannsynligheten for å få en bestemt jobb varierer imidlertid etter hva slags utdannelse aktøren har. Aktøren kan benytte seg av flere metoder for å skille mellom utdannelsene. En mulighet er å plukke ut en jobb og velge den utdannelsen som innebærer størst sannsynlighet for å få den. En annen mulighet er å velge utdannelsen som innebærer høyest sannsynlighet for et veid spekter av jobber.

Det mest kjente bidraget til en probabilistisk valghandlingsmodell for lotterier kom fra Becker et.al.(1963a,b). Det ga støtet til et oppsving i forskningen. Siden den gang har det imidlertid vært gjort svært små fremskritt. To arbeider i den senere tid fra Harless og Camerer (1994) og Hey og Orme (1994) har imidlertid igjen brakt denne typen modeller opp til diskusjon. De sistnevnte tok utgangspunkt i deterministiske modeller for valg mellom lotterier men konkluderer i sitt arbeid med at

> **"Our results suggests quite strongly that the truth is not going to be find along this deterministic choice route, unless some account is taken of the errors. There is clearly a problem of identifying the underlying "true" model because of these errors - indeed it could be argued that the lack of significance for some of the top level functionals (deterministic non - expected utility functionals) for some of the subjects in our study could simply result from this noise,....."**

Dagsvik (1995) tar sikte på å generalisere og Becker et.al.(1963b).

3.1 DETERMINISTISK OG STOKASTISK KOMPONENT

I forbindelse med valg med sikre utfall oppstår det en inkonsistens mellom modellens spesifiserte determinsistiske nyttekomponent og aktørens sanne nytte på grunn av at forskeren har manglende informasjon. Alternativt argumenterte vi med at hverken forskeren eller aktøren selv var i stand til å konsistent evaluere hvert alternativ. I begge tilfellene var det fornuftig spesifisere en stokastisk nytte.

I den nye situasjonen gjelder den samme argumentasjonen. Dersom vi nøyaktig skulle følge argumentasjonen for modeller for sikre utfall, ville det være mest naturlig å definere en stokastisk nytte knyttet til hvert enkelt utfall. Teoretisk er det mulig. Dagsvik (1995) gjør dette i sitt arbeid. Alternativet er å argumentere for en stokastisk nytte knyttet til hvert lotteri. Resultatet av disse to angrepsmåtene er det samme i den forstand at de resulterer i den samme modellen. Jeg vil i denne oppgaven velge den siste fordi den er enklere og mer oversiktlig. Analogt med tidligere er forskeren/aktøren ikke i stand til fullstendig å evaluere nytten $\tilde{V}_i$ knyttet til et lotteri i. Forskeren spesifiserer en deterministisk nyttefunksjon for hvert lotteri, $\tilde{v}_i$. Han innser imidlertid at den avviker fra aktørens sanne. Tidligere ble **uobserverte komponenter i forbindelse med et alternativ** modellert som en stokastisk variabel som inngikk additivt eller multiplikativt sammen med den deteministiske delen. I modeller for lotterier finnes **uobserverte komponenter i forbindelse med utfallene**. De er imidlertid kun en del av bakgrunnen for at nytten knyttet til et lotteri spesifiseres som en stokastisk størrelse. Forskeren vet ikke sikkert hvilken metode aktøren bruker for å vurdere lotteriene. Det er usikkert hvordan aktøren velger å kombinere all informasjon han står ovenfor ved et valg. Denne informasjonen består av utfallsnyttene og utfallssannsynlighetene. Et eksempel jeg skisserte innledningsvis, var valg mellom ulike utdannelser. Forskeren vet ikke om aktøren kun vurderer sannsynligheten for å få den jobben han ønsker seg mest og velger utdannelse ut fra den, eller om han bruker all tilgjengelig informasjon og velger den utdannelsen som gir høyest forventet nytte. Her ligger noe av årsaken til at modeller for lotterier er så vanskelig å formulere. Usikkerheten i en modell blir ofte svært stor. Dersom vi allikevel benytter oss av en argumentasjon analog til den for modeller for valg med sikre utfall, kan nytteverdien til et lotteri spesifiseres som

$$\tilde{V}_i = \tilde{v}_i \tilde{\varepsilon}_i \tag{3.1}$$

bestående av en deterministisk, $\tilde{v}_i$, og en stokastisk del, $\tilde{\varepsilon}_i$.

I forbindelse med modeller for valg med sikre utfall antok vi at nytten var separabel, og den kunne skrives som en lineær funksjon delt opp i to ledd, et deterministisk og et stokastisk. Dette fordi optimeringsproblemene

$$\max_i(U_i) = \max_i(u_i \varepsilon_i)$$

og

$$\max_i(F(U_i)) = \max_i(F(u_i \varepsilon_i))$$

gir samme løsning dersom F er en monoton transformasjon. Følgelig kan vi uten tap av generalitet velge å spesifisere nytten på lineær form selv om det ikke finnes noen teoretisk begrunnelse. Når det gjelder modeller for lotterier, eksisterer det en begrunnelse for lineæritet i form av antagelsen om at aktøren maksimerer forventet nytte. Det er imidlertid ikke sikkert at aktøren bruker denne vurderingsmetoden. Det finnes alternative vurderingsmetoder som det er mulig å argumentere for. Generelt kan nytten knyttet til alternativ i spesifiseres som

$$V_i = \ln(\tilde{V}_i) = v_i + \varepsilon_i = \alpha(u_{iJ}, g_{iJ}) + \varepsilon_i \tag{3.2}$$

hvor u_{iJ} og g_{iJ} beskriver henholdsvis deterministisk utfallsnytte og utfallssannsynlighet og v_i og ε_i beskriver den deterministiske og stokastiske komponenten nytten knyttet til et lotteri. Dersom aktøren maksimerer forventet nytte, kan funksjonen α spesifiseres til å være en lineær kombinasjon av sine argumenter. Tyngdepunktet i min oppgave vil ligge i å finne en fornuftig formulering av α både på et teoretisk og empirisk grunnlag. Jeg vil bruke resten av kapittel 3 på å utdype og problematisere mulige teoretiske formuleringer. Jeg vil så teste disse formuleringene på et empirisk datamateriale. Jeg vil anta at restleddene er uavhengig ekstremfordelte. Det vil si at modellen blir analog til Logit-modellen for sikre utfall (avsnitt 2.1.2.2).

3.2 SPESIFIKASJON AV VURDERINGSMETODE

Hvert lotteri består av en mengde gevinster og tilhørende gevinstsannsynligheter. Nytten knyttet til et lotteri er derfor avhengig av nytten knyttet til hver gevinst og sannsynligheten for å få denne gevinsten. Den deterministiske komponenten kan følgelig skrives som

$$v_i = \alpha(u_{iJ}, g_{iJ}). \tag{3.3}$$

Prinsipielt kan (3.3) sees på som en generalisering av den deterministiske delen i modeller for valg med sikre utfall dersom g_i karakteriseres som et ekstra sett med attributter. I sammenheng med modeller for lotterier har imidlertid α, i motsetning til tidligere, en tolkning. Den beskriver hvordan aktøren kombinerer de to vektorsettene når han vurderer lotteriet. Sagt på en annen måte beskriver den aktørens vurderingsmetode. I denne oppgaven har jeg valgt å konsentrere meg om å spesifisere α. Det vil si at jeg vil fokusere på finne en god spesifikasjon for den vurderingsmetoden det er trolig at aktøren følger.

Jeg vil gjøre tre forenklinger. For det første vil jeg bare se på tilfeller der **hverken attributter eller aktørvariable observeres**. Utfallsnyttens deterministiske del består dermed kun av en utfallsspesifikk konstant, nemlig nytteverdien selv. Videre vil jeg anta at **utfallsnytten er en deterministisk størrelse** eller fullstendig kjent for aktøren. Forskeren måler ikke utfallsnytten direkte. I modeller for valg med sikre utfall ble nyttefunksjonen beskrevet av et sett parametre, β. Forskeren var ikke i stand til å måle disse direkte. Han estimerte dem i stedet på grunnlag av et datamateriale.

Tilsvarende kan vi se på utfallsnyttene som et **estimat** for den nytten en aktør virkelig opplever av et utfall. Utfallsnytten vil dermed inngå som en deterministisk størrelse i modellen selv om den eksakte utfallsnytten er ukjent for forskeren. For å forenkle situasjonen ytterligere vil jeg kun behandle et spesialtilfelle der alle lotteriene består **av samme mengde mulige utfall**. Det vil si at

$$u_J = u_{iJ} \, .$$

Det er flere som har sett på modeller for den typen situasjoner jeg har skissert her. Luce et.al. (s 359-360,1965) beskrev en såkalt Luce-modell for lotterier der han baserer seg på forventningsverdier. Valgsannsynligheten i en Luce modell kan skrives som

$$P(i;C) = \frac{\sum_{j=1}^{K} g_{ij} u_j}{\sum_{t=1}^{S} \sum_{j=1}^{K} g_{tj} u_j} \tag{3.4}$$

der u_J kan tolkes som nytte knyttet til gevinst J. Modellen svarer til å spesifisere nytten knyttet til lotteri i gitt ved

$$V_i = E_{g_i}(U_J) = \sum_{j=1}^{K} g_{ij} u_j + \varepsilon_i \tag{3.5}$$

der ε_i , i=1,2,..,S, er uavhengig og identisk ekstremfordelte størrelser. Modellen beskriver følgelig en aktør som velger det lotteriet som maksimerer forventet nytte i tillegg til et støyledd.

Becker et.al. (1963b) presenterte en mer generell modell. Den inneholdt modell (3.4) som et spesialtilfelle. Valgsannsynligheten var gitt ved

$$P(i;C) = \frac{h\left(\sum_{j=1}^{K} g_{ij} u_j\right)}{\sum_{t=1}^{S} h\left(\sum_{j=1}^{K} g_{tj} u_j\right)} \tag{3.6}$$

der h er en vilkårlig monoton transformasjon. Dersom h er identitetstransformasjonen, er modellen identisk med Luce' modell. Modellen (3.6) innebærer tilsvarende som (3.5) en spesifikasjon av nytten knyttet til lotteri i gitt ved

$$V_i = \ln\left(h\left(E_{g_i}(u_J)\right)\right) + \varepsilon_i = \ln\left(h\left(\sum_{j=1}^{K} g_{ij} u_j\right)\right) + \varepsilon_i \; . \tag{3.7}$$

Denne modellen er ikke testbar siden det ikke er gjort en mer spesifikk antagelse av h. Becker et.al. tester i sitt arbeid kun om den er konkav, lineær eller konveks. Han fant ut at det så ut som om formen varierte med aktørene. Hansen (1995) gjennomførte den samme undersøkelsen basert på andre data og kom frem til samme konklusjon.

Dersom vi tar utgangspunkt i at aktøren vurderer et lotteri ut fra forventet nytte og støy, er det ikke gitt at det er Luce' modell som beskriver denne vurderingsmetoden på en best mulig måte. Det finnes flere spesifiseringer av h i tråd med denne antagelsen. Vi er interessert i å finne den som beskriver aktøren best. Det er ikke mulig å argumentere for en spesiell h ut fra teorien. Det gjør det nødvendig å basere et eventuelt valg av h på en empirisk undersøkelse. Jeg gjennomfører en slik empirisk estimering i kapittel 5. For å operasjonalisere Becker's modell, vil jeg betrakte en spesiell transformasjon, h, som inneholder Luce' modell som et spesialtilfelle. Den spesielle transformasjonen resulterer i en valgsannsynlighet gitt ved

$$P(i;C)=\frac{\left(\sum_{j=1}^{K} u_j^{\theta} g_{ij}\right)^{\frac{1}{\theta}}}{\sum_{t=1}^{S}\left(\sum_{j=1}^{K} u_j^{\theta} g_{tj}\right)^{\frac{1}{\theta}}} \tag{3.8}$$

det vil si at V_i kan utrykkes ved

$$V_i = \tfrac{1}{\theta}\ln\left(E_{g_i}\left(u_j^{\theta}\right)\right)+\varepsilon_i = \tfrac{1}{\theta}\ln\left(\sum_{j=1}^{K} u_j^{\theta} g_{ij}\right)+\varepsilon_i \tag{3.9}$$

der θ er en positiv parameter som gir ulike monotone transformasjoner.

3.3 TOLKNING AV θ

Aktøren bruker en vurderingsmetode for å velge mellom lotteriene. Problemet når denne vurderingsmetoden skal spesifiseres i en modell, er at det ikke er opplagt i enhver situasjon hvilken vurderingsmetode aktøren bruker. Jeg presenterte i den forbindelse en modell som var bestemt på en parameter nær; nemlig θ. Avhengig av denne parameteren er ulike vurderingsmetoder spesifisert i modellen. Jeg vil i det følgende se på **sammenhengen mellom θ og aktørens vurderingsmetode**.

Jeg vil i den forbindelse ta for meg to utgaver av modellen i (3.8); en slik den er spesifisert til nå og en forenklet utgave. Jeg vil starte med den siste. Vi har foreløpig hverken spesifisert verken form eller nivå av utfallsnytten. Det medfører derfor ikke noe tap av generalitet dersom vi substituerer $\beta_j = u_j^{\theta}$ inn i (3.9):

$$V_i = \tfrac{1}{\theta}\ln\left(\sum_{j=1}^{K} \beta_j g_{ij}\right)+\varepsilon_i. \tag{3.10}$$

Nytten er følgelig definert for $0 < \theta < \infty$. Modellen endres ikke dersom nytten spesifiseres som

$$\overline{V}_i = V_i\theta = \ln\left(\sum_{j=1}^{K} \beta_j g_{ij}\right)+\theta\varepsilon_i. \tag{3.11}$$

Følgelig bestemmer θ, sammen med nyttenivået, stigningen av fordelingsfunksjonen. Det er vanlig å anta at aktøren vurderer et lotteri utfra forventet nytte. **I så måte viser θ i hvilken grad aktøren velger konsistent**. Han velger konsistent dersom han velger det lotteriet som det er knyttet høyest forventet nytte til. Ved høye θ-verdier relativt til nyttenivået er ikke aktøren i stand til å evaluere forventnings-verdien, og valget er styrt av tilfeldige mekansimer. Denne argumentasjonen forutsetter at V_i er stokastisk også for aktøren. Dersom ε_i kun er stokastisk for forskeren, **viser θ i hvilken grad forskeren har klart å beskrive aktørens preferanser**. Det vil si at θ viser i hvor stor grad aktøren vurderer et lotteri på grunnlag av den naturlige logaritmen til den forventede nytten som forskeren har definert. I tilfellet $\theta = 1$ reduseres modellen til en Luce-modell (3.4) der sannsynligheten for å velge lotteri i er proporsjonal med forventet nytte knyttet til lotteri i.

Til nå har jeg sett på modeller med parametre i intervallet < 0 , ∞ >. Jeg vil så gå over til å undersøke grensetilfellene $\theta = 0$ og $\theta = \infty$. For at det numerisk skal være mulig å estimere slike modeller, må vi ta utgangspunkt i den utgaven som er spesifisert i (3.9). Først vil jeg ta for meg tilfellet der $\theta = 0$. Det er mulig å finne grenseverdien

$$\lim_{\theta\to 0} V_i = \lim_{\theta\to 0}\left(\tfrac{1}{\theta}\ln\left(\sum_{j=1}^{K} u_j^{\theta} g_{ij}\right) + \varepsilon_i\right) \tag{3.12}$$

ved hjelp av L'Hôpital's regel. Derivasjon av teller og nevner gir

$$\lim_{\theta\to 0} V_i = \lim_{\theta\to 0}\left(\frac{\sum_{j=1}^{K} u_j^{\theta} g_{ij}\ln(u_j)}{\sum_{j=1}^{K} u_j^{\theta} g_{ij}}\right) + \varepsilon_i = \sum_{j=1}^{K} g_{ij}\ln(u_j) + \varepsilon_i = E_{g_i}(\ln(u_j)) + \varepsilon_i \tag{3.13}$$

som resulterer i valgsannsynligheten

$$P(i;C) = \frac{e^{\sum_{j=1}^{K} g_{ij}\ln(u_j)}}{\sum_{t=1}^{S} e^{\sum_{j=1}^{K} g_{tj}\ln(u_j)}}. \tag{3.14}$$

Den resulterende modellen har en deterministisk del som i likhet med Luce' modell beskriver en aktør som i tillegg til et støyledd vurderer lotteriene utfra deres forventede nytte. Da det finnes to spesifikasjoner av denne vurderingsmetoden, kan det være interessant å undersøke hvilken som beskriver den best. Vi har tidligere argumentert for at en transformasjon av utfallsnytten ikke medfører tap av generalitet. Det innebærer at vi kan sette $\delta_J = \ln(u_J)$ inn i (3.13), noe som gir et bedre grunnlag for å sammenligne og tolke forskjellen mellom de to modellene med $\theta = 0$ og $\theta = 1$. For den sistnevnte modellen gjelder følgende ekvivalens

$$\max_i\left(\ln\left(\sum_{j=1}^{K}\beta_j g_{ij}\right) + \varepsilon_i\right) \Leftrightarrow \max_i\left(\left(\sum_{j=1}^{K}\beta_j g_{ij}\right)e^{\varepsilon_i}\right) \tag{3.15}$$

som vi skal sammenligne med

$$\max_i\left(\left(\sum_{j=1}^{K}\delta_j g_{ij}\right) + \varepsilon_i\right). \tag{3.16}$$

Av dette ser vi at modellen med parameter $\theta = 1$ (3.15) innebærer et multiplikativt avvik i forhold til forventet nytte, mens modellen med $\theta = 0$ (3.16) innebærer et additivt avvik. Til nå er det så vidt jeg vet ingen som har klart å argumentere på et teoretisk grunnlag for at den ene modellen kan foretrekkes framfor den andre. Det er mulig å trekke en parallell til modeller for sikre utfall. I en del situasjoner ga transformasjon av en variabel en bedre modell (avsnitt 2.1.1). Den eneste måten for å konstatere om en modell med en transformert variabel, ble en bedre modell, var ved hjelp av empiri. **Siden vi her ikke er i stand til å skille mellom modellene på teoretisk grunnlag, må vi også i denne forbindelse ty til eksperimentdata for å bestemme hvilken av modellene som er den beste**. Jeg gjør et forsøk på dette i kapittel 5.

Til slutt vil jeg se på tilfellet der $\theta \to \infty$. Jeg bruker som før L'Hôpital's regel for å finne grenseverdien

$$\lim_{\theta\to\infty} V_i = \lim_{\theta\to\infty}\left(\frac{\sum_{j=1}^{K} u_j^{\theta} g_{ij} \ln(u_j)}{\sum_{j=1}^{K} u_j^{\theta} g_{ij}}\right) + \varepsilon_i. \tag{3.17}$$

Jeg dividerer så med den høyeste nytteverdien i teller og nevner av (3.21). Resultatet

$$\lim_{\theta\to\infty} V_i = \lim_{\theta\to\infty}\left(\frac{\sum_{j=1}^{K}\left(\frac{u_j}{\max_{t\in D}(u_t)}\right)^{\theta} g_{ij}\ln(u_j)}{\sum_{j=1}^{K}\left(\frac{u_j}{\max_{t\in D}(u_t)}\right)^{\theta} g_{ij}}\right) + \varepsilon_i = \ln\left(\max_{j\in D}(u_j)\right) + \varepsilon_i \tag{3.18}$$

viser at **aktøren kun evaluerer nytteverdien av et utfall, ikke sannsynligheten for det**. Med de forenklingene jeg har gjort (kapittel 3.2), har alle lotteriene de samme gevinstene. Optimeringsproblemet som følger av $\theta\to\infty$ er dermed gitt ved

$$\max_i V_i\Big|_{\theta\to\infty} = \max_i\left(\ln\left(\max_{j\in D}(u_j)\right) + \varepsilon_i\right) = \ln\left(\max_{j\in D}(u_j)\right) + \max_i(\varepsilon_i). \tag{3.19}$$

Det første leddet i (3.19) er konstant. Valget avhenger følgelig kun av andre ledd og blir helt stokastisk. Valgsannsynligheten er dermed gitt ved

$$P(i;C) = \frac{1}{K}$$

der K er antall valgbare lotterier. I vårt forenklede tilfelle med lotterier bestående av samme utfallsmengde, klarer ikke aktøren å skille mellom lotteriene ved hjelp av denne vurderingsmetoden. Grensetilfellet for $\theta\to\infty$ i (3.18) står ikke i kontrast til tolkningen jeg gjorde på grunnlag av (3.11). Utfra (3.11) var θ et mål for hvor konsistent aktøren valgte, alternativt hvor godt aktørens valghandling var beskrevet av forskeren. En modell med en stor θ indikerte at valget til en stor grad var styrt av tilfeldige mekanismer. Modellen i (3.18) beskriver således det mest ekstreme tilfellet der valget er helt tilfeldig.

Dersom gevinstene imidlertid er ulike fra lotteri til lotteri, beskriver (3.18) en aktør som kun evaluerer den informasjon som inngår i nytteverdien knyttet til hvert utfall. Årsaken til at en aktør velger slik vurderingsmetode kan være at han står ovenfor så mye informasjon, at han ikke er i stand til å vurdere alt. Vurderingsmetoden kan således sees på som en forenkling av en komplisert valgsituasjon. Alternativt betyr ikke sannsynlighetsfordelingen noe for aktøren. Han er villig til å inngå en større risiko det vil si godta en mindre sannsynlighet, for å få den høyest vurderte gevinsten.

3.4 ALTERNATIV MODELL

Jeg nevnte i avsnitt 3.1 at dersom man ville utlede en modell for lotterier helt i analogi med en modell for sikre utfall, hadde det naturlige vært å spesifisere en stokastisk nytte knyttet til hvert enkelt utfall. Det en imidlertid en del regnetekniske problemer forbundet med en slik fremgangsmåte. Jeg valgte derfor en annen tilnærming i avsnitt 3.1. Det finnes imidlertid spesialtilfeller der en modell kan utledes fra stokastisk utfallsnytte på en forholdsvis enkel og elegant måte. For å illustrere at det er mulig, vil jeg ta for meg et slikt tilfelle her.

La oss betrakte en situasjon der hvert lotteri har et **stort antall mulige utfall**. I analogi med en modell for sikre utfall, spesifiserer jeg en separabel og additiv nyttefunksjon

$$U_J = u_J + \eta_J. \tag{3.20}$$

Restleddet η_J kan være et resultat av at forskeren ikke er i stand til å observere alle de komponentene som inngår i aktørens nyttefunksjon. Jeg antar så at aktøren vurderer et lotteri ut fra forventet nytte. Denne nytten er gitt ved

$$V_i = E_{g_i}(U_J) = \sum_{j=1}^{K} g_{ij}\, u_j + \sum_{j=1}^{K} g_{ij}\, \eta_j \tag{3.21}$$

Antar at videre at η_J er uavhengig fordelt med

$$E(\eta_J) = 0$$

$$Var(\eta_J) = \sigma^2$$

noe som medfører at

$$E\left(E_{g_i}(\eta_J)\right) = 0$$

$$Var\left(E_{g_i}(\eta_J)\right) = \sum_{j=1}^{K} g_{ij}^2 \sigma^2.$$

Siden antall mulige utfall i hvert lotteri er svært stort, vil Sentralgrenseteoremet gjelde. Det vil si at $Y_i = E_g(\eta_J)$ er tilnærmet normalfordelt. Normalfordelte restledd resulterer i en valghandlingsmodell av type Probit (avsnitt 2.1.2.1). Ved flere enn to lotterier innebærer denne svært kompliserte algoritmer for estimering. Siden jeg kun er interessert i å illustrere et prinsipp, vil jeg konsentrere meg om det binære tilfellet.

Valgsannsynligheten i en modell med to lotterier er gitt ved

$$P(1;\{1,2\}) = \left(\sum_{j=1}^{K} g_{1j}v_j + Y_1 > \sum_{j=1}^{K} g_{2j}v_j + Y_2\right) = P\left(Y_2 - Y_1 < \sum_{j=1}^{K} g_{1j}v_j - \sum_{j=1}^{K} g_{2j}v_j\right) \tag{3.22}$$

hvor $Y = Y_2 - Y_1$ er normalfordelt med

$$E(Y) = 0$$

$$\mathrm{Var}(Y) = \left(\sum_{j=1}^{K} g_{1j}^2 + \sum_{j=1}^{K} g_{2j}^2\right)\sigma^2.$$

Følgelig blir

$$P(1;\{1,2\}) = \Phi\left(\frac{\sum_{j=1}^{K} g_{1j}u_j - \sum_{j=1}^{K} g_{2j}u_j}{\sigma\sqrt{\left(\sum_{j=1}^{K} g_{1j}^2 + \sum_{j=1}^{K} g_{2j}^2\right)}}\right) \tag{3.23}$$

hvor Φ er den standardiserte, kumulative normalfordelingen.

4. ESTIMERING

Frem til nå har jeg kun sett på spesifisering av modeller på en teoretisk basis. Disse modellene inneholder en mengde ukjente parametre knyttet til hver variabel. I modeller for valg med sikre utfall betegnet jeg parametervektoren med $\boldsymbol{\beta}_n$ og i modeller for lotterier med **u**. Så langt er imidlertid disse parameterene ukjente for forskeren. Estimeringsprosessen består i å bestemme dem på grunnlag av et datamateriale. Parameterverdiene som bestemmes i en slik prosess, kalles for estimater. Den sanne parameterverdien er imidlertid ukjent, og det er viktig å være klar over at estimatet ikke nødvendigvis stemmer overens med denne. Årsaken er til dels at estimatet avhenger av datamaterialet eller egentlig den delen av populasjonen det er samlet inn fra, det såkalte samplet. Dette avhengighetsforholdet gjør at det er nødvendig å sette en del krav til datamaterialet for å få gode estimater og modeller. Jeg vil gå nærmere inn på dette i kapittel 4.1. Videre kan estimatet avvike fra den sanne parameterverdien fordi modellen beskriver en heterogen populasjon. Det vil si at aktørene har ulike preferanser som modellen ikke får fanget opp. Estimatene blir følgelig en "kompromiss"- løsning. En modell er i de fleste tilfeller ikke entydig definert. Det å spesifisere en modell blir derfor en iterativ prosess. I denne prosessen er det et behov for å vurdere estimater og sammenligne alternative modeller. Jeg vil i avsnitt 4.2 ta for meg de viktigste vurderings -og sammenligningsmetodene. I kapittel 4.3 beskriver jeg estimeringsmetoden.

4.1 KRAV TIL DATAMATERIALET

Som jeg nevnte innledningsvis, er kvaliteten av estimatene i en modell avhengig av datamaterialet. Det er derfor i innsamlingsfasen viktig å ta hensyn til en del krav. Jeg vil i det følgende gå inn på to svært viktige krav for å få gode estimater.

4.1.1 Balansert utfallsrom

Det må være et **minste antall utfall for hvert valgalternativ eller lotteri**, for at det i det hele tatt skal være mulig å foreta en estimering. Sagt på en annen måte, må et minimum av aktørene ha valgt hvert enkelt alternativ eller lotteri. Det finnes to måter å få dette til på. Den mest opplagte er å observere en så stor gruppe aktører at for eksempel minimum fem har valgt hvert element i valgmengden. Sannsynligheten for at et alternativ eller lotteri velges er alltid positiv, slik at antall som har valgt et alternativ eller lotteri øker monotont med antall spurte. I mange tilfeller har forskeren mulighet til å være med på å utforme alternativene eller lotteriene. Det er da viktig å utforme alternativene eller lotteriene slik at utfallsrommet blir balansert. Det vil si at de har en noenlunde lik attraktivitet for størsteparten av de spurte.

4.1.2 Spredning av variable

Jeg vil som illustrasjon trekke en parallell til tradisjonell regresjon. Ved lineær regresjon estimeres en linje. For å få en god modell bør **variablene være spredd over hele regresjonsintervallet**. Dette gjelder også for Logit -og Probitmodeller. Dersom et attributt ikke varierer fra utfall til utfall i lotteriene eller fra alternativ til alternativ i en modell med sikre utfall, gir de ikke modellen økt forklaringsgrunnlag. Et eksempel på dette er på restaurant om prisen for de ulike rettene er nesten lik. Å bruke pris som forklaringsvariabel, tilsvarer da bare å legge til en konstant i hver enkelt nyttefunksjon. Dette endrer som forklart før, ikke modellen, og attributtet kan utelates.

4.2 METODER FOR VURDERING AV MODELLEN

For å vurdere om en modell er god, finnes det en del kjente metoder. Jeg vil i dette avsnittet gå igjennom de mest relevante i denne sammenheng.

En relevant problemstilling når en modell skal spesifiseres, er å finne ut **hvilke variable som har betydning i modellen**. En variabel har betydning i modellen dersom den tilhørende parameteren er forskjellig fra null. Forskeren har imidlertid ikke kjennskap til den nøyaktige parameterverdien, han kan kun finne et estimat for den. Dette estimatet kan være forskjellig fra null selv om variabelen ikke har noen innvirkning. Vi sier derfor at et estimat må være signifikant forskjellig fra null før den tilhørende variabelen kan sies å ha betydning. Det finnes flere måter for å teste om et estimat er signifikant forskjellig fra null. En **t-test** er et eksempel. Jeg vil ikke gå nærmere inn på denne her og viser i stedet til Lerman og Ben-Akiva (1985, s.23-26). Spesifisering av en modell er en iterativ prosess. Forskeren spesifiserer i første omgang en modell. Så estimeres parametrene i modellen. Ut fra resultater av t-tester for alle parametrene, må modellen omformuleres ved å ta ut eller legge til forklaringsvariable.

En annen fremgangsmåte for å teste en modell er å legge til en eller flere beskrankninger og sammenligne den opprinnelige modellen med den beskrankede. En beskrankning kan eksempelvis være å sette en parameter lik null. Problemet vil da tilsvare en t-test. Det finnes imidlertid også en rekke andre mulige beskrankninger. En som er spesielt aktuell i denne forbindelse, er å tilegne θ en spesiell verdi i en modell for lotterier. Det er da mulig å teste om en spesiell fiksering av θ gir en signifikant dårligere modell. Vi kan teste dette ved hjelp av en sannsynlighetskvote-eller **LR-test**. "Likelihood"-verdien L som vi skal komme mer tilbake til i avsnitt 4.3, er en simultansannsynlighet for alle observasjonene i observasjonsmaterialet, innsatt estimerte parametre og variabelverdiene. Dersom θ i (3.8) settes til en bestemt verdi k, tester LR-testen nullhypotesen

$$H_0 : \theta = k \qquad \text{mot} \qquad H_1 : \theta \neq k$$

ved hjelp av testobservatoren

$$LR = -2\left(\ln\left(L\left(\hat{\theta}\right)\right) - \ln(L(k))\right)$$

der $\hat{\theta}$ er estimat for θ i en modell for lotterier (3.8). Dersom nullhypotesen er sann, er LR χ^2-fordelt med et antall frihetsgrader, *v*, svarende til antall beskrankninger.

Som beskrevet i avsnitt 3.1.1 hadde det optimale vært å estimere en individuell parametervektor. Det er av regnetekniske årsaker ikke mulig da antall observasjoner er for lite. En løsning på problemet er å estimere en felles parametervektor for alle aktørene i samplet. Spørsmålet er imidlertid om aktørene danner en homogen gruppe, slik at en slik felles beskrivelse er fornuftig. Det er mulig å teste dette ved å se på om en oppdeling i T segment gir signifikant bedre modeller, noe som tilsvarer nullhypotesene

$$H_0 : \beta = \beta_t \qquad \text{mot} \qquad H_1 : \beta \neq \beta_t \qquad \text{for} \qquad t = 1,2,\ldots,T$$

der β er parametre i modellen uten segmentering, og β_t parametre i en modell for segment t.

Nullhypotesen testes av testobservatoren

$$LR = -2\left(\ln\left(L\left(\hat{\beta}\right)\right) - \sum_{t=1}^{T} \ln\left(L\left(\hat{\beta}_t\right)\right)\right)$$

som er χ^2-fordelt med

$$v = \sum_{t=1}^{T} K_t - K$$

frihetsgrader. K er antall parametre i modellen for hele observasjonsmaterialet, mens K_t er antall parametre i modell for segment t.

Det finnes flere metoder for å sammenligne flere modeller, tar utgangspunkt i i hvilken grad en modell beskriver observasjonsmaterialet, såkalt **"goodness of fit"**. I forbindelse med tradisjonell regresjonsanalyse er "goodness of fit" gitt ved R^2 eller justert-R^2. I forbindelse med kvalitative valg-handlingsmodeller finnes et lignende mål. Det er gitt ved

$$\rho^2 = 1 - \frac{\ln\left(L\left(\hat{\beta}\right)\right)}{\ln(L(0))}. \tag{4.1}$$

Et problem er at, som R^2, øker ρ^2 alltid med antall variable i modellen. Det vil si at dersom en ny variabel tillempes modellen, vil ρ^2 alltid indikere at modellen blir bedre. Dette kompenseres for i et justert mål

$$\bar{\rho}^2 = 1 - \frac{\ln\left(L\left(\hat{\beta}\right)\right) - K}{\ln(L(0))} \tag{4.2}$$

der K tilsvarer antall parametre i modellen. Modellen som forklarer observasjonsmaterialet best, er den med høyest $\bar{\rho}^2$.

Et videre mål for å sammenligne modeller er **prediksjonsfeil** gitt ved

$$\bar{E} = \sum_{n=1}^{N} \frac{1 - P_n(i_n)}{N} \tag{4.3}$$

der i_n er alternativ eller lotteri valgt av aktør n. Prediksjonsfeilen viser i likhet med "goodness og fit" i hvilken grad modellen beskriver datamaterialet, eller egentlig i hvilken grad den ikke beskriver det. Den beste modellen har den minste prediksjonsfeilen.

4.3 SANNSYNLIGHETSMAKSIMERING (SM)

Dersom en foreløbig modell er formulert og datamaterialet samlet inn, kan estimeringen starte. En vanlig fremgangsmåte for å finne ukjente parametere er å løse et ligningssystem. Generelt for ligningsystemer gjelder at de har eksakt en løsning hvis og bare hvis antall uavhengige ligninger svarer til antall ukjente. I forbindelse med estimering vil det imidlertid nesten alltid være slik at antall uavhengige ligninger er større enn antall ukjente. Ligningssystemet kalles i slike tilfeller for over-bestemt, og det finnes flere metoder for å finne en tilnærmet løsning. I forbindelse med kvalitative valghandlingsmodeller er sannsynlighetsmaksimeringsmetoden den mest brukte. Jeg vil derfor kun gå inn på denne her. Utgangspunktet for all estimering, er et datamateriale med N observasjoner. For hver observasjon finnes S vektorer med variabelverdier, en svarende til hvert alternativ eller lotteri. SM-metoden bestemmer et parameterestimat som **maksimerer simultan sannsynlighet L** for de N observasjonene, innsatt alle variabelverdier. Ved estimering av en modell velges parametervekoren $\hat{\mathbf{u}}$ dersom

$$L\left(\hat{\mathbf{u}}\right) = \max_{\mathbf{u}}\left(\prod_{n=1}^{N}\prod_{i=1}^{S}\left(P_n(i;C)\right)^{Y_{in}}\right) = \max_{\mathbf{u}}\left(\prod_{n=1}^{N}\prod_{i=1}^{S}\left(\frac{e^{V_i(\mathbf{u})}}{\sum_{t=1}^{S} e^{V_i(\mathbf{u})}}\right)^{Y_{in}}\right) \tag{4.4}$$

hvor

$$Y_{in} = \begin{cases} 1 \text{ dersom lotteri i velges i observasjon n} \\ 0 \text{ ellers} \end{cases}$$

og L er "likelihood"-funksjonen.

McFadden (1974) viste at likelihoodsfunksjonen har et globalt maksimum når den deterministiske delen av nyttefunksjonen er lineær i sine parametre. I den modellen vi definerte i (3.9) var den deterministiske delen ulineær. Vi er følgelig ikke garantert et entydig maksimalpunkt.

Regneteknisk gjenstår nå et problem. Det består i at ligningssystemet ikke er lukket løsbart. Parameterene må dermed beregnes iterativt.

5. REISEMÅLSEKSPERIMENT

5.1 FORMÅL

I kapittel 3 skisserte jeg en modell for lotterier. Denne modellen var en spesialutgave av modellen til Becker et.al. (3.6). Modellen varierer med en parameter θ. Denne parameteren kan anta verdier slik at modellen tilsvarer Luce modell (3.4) eller modeller med lignende forventet-nytte vurderinger. Hvilken modell som beskriver valghandlingen best, er imidlertid uvisst. Becker et.al. (1963b) konkluderte med at dette varierte fra aktør til aktør. Jeg vil i dette eksperimentet estimere særtilfeller av modellen i (3.8) for ulike verdier av θ for å finne ut hvilken modell som best gjenskaper observert valgadferd.

5.2 BESKRIVELSE AV EKSPERIMENT

Undersøkelsen består av en rekke konstruerte lotterier. I hvert av lotteriene er gevinsten **en reise**. Denne gevinsten blir trukket ut fra en mengde bestående av en eller flere av de fire reisene listet opp nedenfor.

- Reise 1 som går til Kairo (Egypt)
- Reise 2 som går til Beijing (Kina)
- Reise 3 som går til Bahamas (Karibien)
- Reise 4 som går til Dublin (Irland)

Det er kun mulig å vinne en reise i hvert lotteri. Sannsynligheten for å vinne en bestemt gevinst i et lotteri er kjent og varierer fra lotteri til lotteri. Et lotteri er eksempelvis definert ved

	Kairo	Beijing	Bahamas	Dublin
Lotteri x	0.65	0.1	0.1	0.15

der sannsynligheten for å vinne de ulike gevinstene er henholdsvis 0.65, 0.1, 0.1 og 0.15. En oversikt over samtlige konstruerte lotterier er gitt i vedlegg D.

Lotteriene grupperes, og aktøren blir bedt om å velge et lotteri i hver av disse undergruppene. Aktøren kan i hver undergruppe eller **eksperiment** velge mellom **tre ulike lotterier**. Hele undersøkelsen består av 116 ulike eksperiment. Det vil si at det finnes totalt 116*3 lotterier. Eksperimentene er ulike i den forstand at lotteriene de består av, inneholder ulike gevinstsannsynligheter. Alle lotteriene tilbyr muligheten for å vinne de samme gevinstene, men sannsynlighetene for å vinne dem varierer. Et bestemt lotteri kan inngå i flere av eksperimentene, men kombinasjonen av lotterier i et eksperiment er unik.

Ni aktører deltok i undersøkelsen. Hver av dem valgte 116 ganger. Det vil si at datamaterialet består av 116*3*9=1044 observasjoner. Disse observasjonene finnes i vedlegg E. Beskrivelsen av eksperimentene som ble presentert for aktørene finnes i vedlegg D. Datamaterialet er samlet inn av Anett Christin Hansen (se Hansen(1995)).

5.3 BESKRIVELSE AV TESTMODELL OG ESTIMERINGS PROSESS

Gevinstsannsynlighetene i r'te eksperiment er gitt ved matrisen

$$G_r = \begin{bmatrix} g_{11r} & g_{12r} & g_{13r} & g_{14r} \\ g_{21r} & g_{22r} & g_{23r} & g_{24r} \\ g_{31r} & g_{32r} & g_{33r} & g_{34r} \end{bmatrix} \tag{5.1}$$

der elementet g_{iJ} beskriver sannsynligheten for å vinne reise J i lotteri i. Matrisen er gitt og kjent både for aktør og meg. Det finnes 116 ulike matriser, se vedlegg D.

Den deterministiske nytten knyttet til gevinst J betegnes med u_J der $j \in \{1,2,3,4\}$; aktøren velger mellom tre lotterier hvor han kan vinne en av fire gevinster. Modellen blir svært enkel. Det eneste som skiller et lotteri fra et annet er gevinstsannsynlighetene. Dermed blir nytteverdiene v_J de parametre som estimeres i modellen.

Jeg vil nå ta for meg den generelle modellen. Nytteverdiene knyttet til hvert lotteri er gitt ved

$$V_i = \tfrac{1}{\theta} \ln\left(\sum_{j=1}^{4} u_j^\theta g_{ij}\right) + \varepsilon_i = v_i + \varepsilon_i \ , \qquad i = 1,2,3. \tag{5.2}$$

Jeg antar så at restleddene er identisk, uavhengig ekstremfordelt slik at sannsynligheten for valg av lotteri 1 er gitt ved

$$P(1;\{1,2,3\}) = P(V_1 > \max(V_2, V_3)) = \frac{e^{v_1}}{\sum_{t=1}^{3} e^{v_t}} \ . \tag{5.3}$$

Modellen er prinsipielt helt analog til en multinomisk Logit-modell for sikre utfall. Valgsannsynligheten generelt gitt ved

$$P(i;C) = \frac{\left(\sum_{j=1}^{K}(cv_j)^\theta g_{ij}\right)^{1/\theta}}{\sum_{t=1}^{S}\left(\sum_{j=1}^{K}(cv_j)^\theta g_{tj}\right)^{1/\theta}} = \frac{(\sum_{j=1}^{K}(v_j)^\theta g_{ij})^{1/\theta}}{\sum_{t=1}^{S}(\sum_{j=1}^{K}(v_j)^\theta g_{tj})^{1/\theta}} . \tag{5.4}$$

Vi ser at det ikke er mulig å bestemme c. I estimeringen medfører dette at vi er nødt til å sette et nivå. Vi står fritt til å velge en nytteverdi knyttet til et av utfallene i lotteriene. Det er vanlig å sette denne nytteverdien lik 1 eller 0 avhengig av hvilken verdi θ antar. Å sette nytteverdien lik null kan gi numeriske problemer i optimeringsprosessen i tilfeller der $\theta = 0$ i forbindelse med ln(0) og 1/0. Jeg velger derfor å sette $u_1=1$ i de tilfeller der θ er en parameter i modellen og $u_1=0$ ellers.

Jeg estimerer parametrene u_2, u_3, u_4 og θ ved hjelp av **sannsynlighetsmaksimeringsmetoden** (avsnitt 4.3). Det vil si at de parametre i modellen som velges, gir maksimal simultan sannsynlighet for observasjonsmaterialet.

Simultan sannsynlighet er gitt ved

$$L = \prod_{n=1}^{1044} \prod_{i=1}^{3} P(i;C)^{Y_{in}} = \prod_{n=1}^{1044} \frac{e^{\sum_{i=1}^{3} Y_{in} v_i}}{\sum_{i=1}^{3} e^{v_i}} \tag{5.5}$$

eller

$$\ln(L) = \sum_{n=1}^{1044} \left[\sum_{i=1}^{3} Y_{in} v_i - \ln\left(\sum_{i=1}^{3} e^{v_i} \right) \right]. \tag{5.6}$$

Optimeringen foregår ved hjelp av et programmeringsverktøy kalt **TSP**, Time Series Prosessor. Som navnet tilsier ble programmet opprinnelig utviklet for tidsserier. Det egner seg imidlertid godt til å analysere de fleste typer repeterte data. Observasjoner og gevinstsannsynligheter blir lagret som serier, der hver serie har et navn. TSP gjør det mulig å operere på en serie ved kun å nevne seriens navn. For videre beskrivelse av programmet viser jeg til "TSP User's Guide" og "TSP Reference Manual". Utdrag fra programkoden finnes kommentert i vedlegg F.

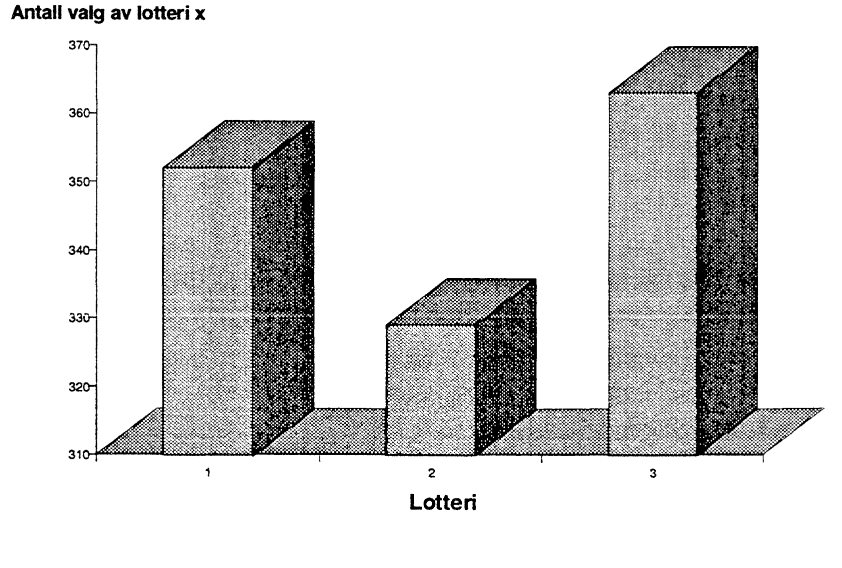
Estimeringen kan mislykkes av flere årsaker. Det kan oppstå numeriske problemer. Siden det inngår flere potenser i målfunksjonen (5.6), kan tallene bli så store at de går utover maskinkapasiteten. Denne kapasiteten er begrenset av e^{88}. Videre inngår ln-funksjoner som kun tar strengt positive argumenter. Ln-transformasjonen av et tall som er negativt eller lik null, eksisterer ikke. Dette kan forårsake numeriske feil i estimeringsprosessen. For særtilfellet $\theta = 0$ må man ta ekstra forholdsregler under estimeringen siden $\frac{1}{0}$ ikke eksisterer. Alternativene blir enten å estimere en modell ut fra (3.13) eller å estimere en modell ut fra (3.8) innsatt en svært liten men positiv θ. Resultatet blir tilnærmet det samme.

5.4 RESULTATER

5.4.1 Oversikt over data

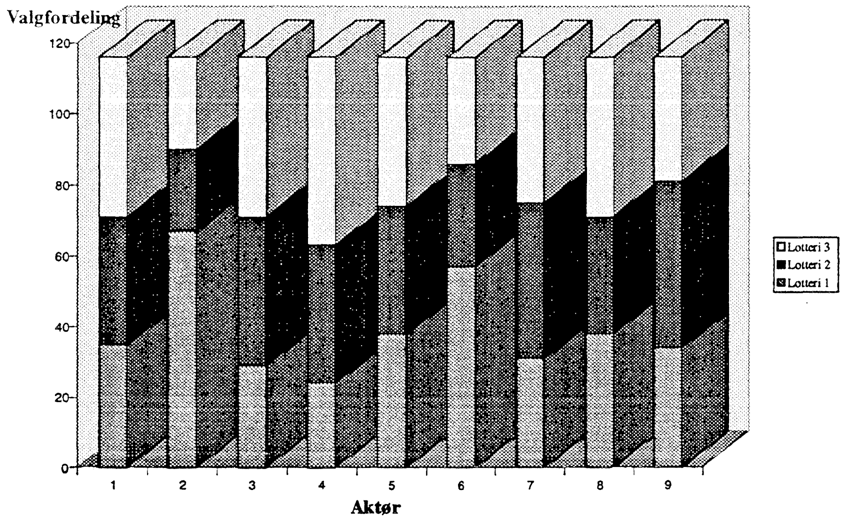
Jeg nevnte i avsnitt 4.1 to viktige kriterier for å få gode estimater. For det første måtte utfallsrommet være balansert. Vi ser av figur 5-1 at dette er oppfylt, dersom vi betrakter hele observasjonsmaterialet samlet.

Figur 5-1: Valgfrekvens for de tre lotteriene



Dersom vi er interessert i å estimere individuelle modeller, bør utfallsrommet være balansert også for den enkelte aktør.

Figur 5-2: Valgfrekvens for de tre lotteriene for hver enkelt aktør.



Figur 5-2 viser for hver aktør hvor mange ganger hvert enkelt lotteri er valgt. Vi ser at fordelingen er forholdsvis balansert. De fleste av aktørene har valgt hver av lotteriene omtrent 39 ganger. Den som avviker mest fra denne fordelingen, er aktør 2. Han har valgt lotteri 1 hele 67 ganger. Det er mulig at det er vanskelig å estimere en modell som beskriver denne aktøren.

Et videre krav for å få gode estimater, er at variablene skal være spredd over hele mulighetsområdet. Dette området er gitt ved $0 \leq g_{ij} \leq 1$. Dersom vi betrakter de konstruerte lotteriene i vedlegg D, ser vi at utfallssannsynlighetene antar verdier i hele intervallet.

5.4.2 Estimering av en modell på grunnlag av hele observasjonsmaterialet

Jeg foretok estimering av fire tilfeller av modellen beskrevet i (5.3). Disse fire vil jeg i det følgende kalle modell A, B, C og D. De skiller seg fra hverandre ved å ha en ulik θ-verdi; henholdsvis -0.87, 0 ,1 og 10. I modell A lot jeg θ variere som en parameter, mens den var fiksert til en spesiell verdi i de andre modellene. Resultatet av estimeringene er vist i tabell 5-1.

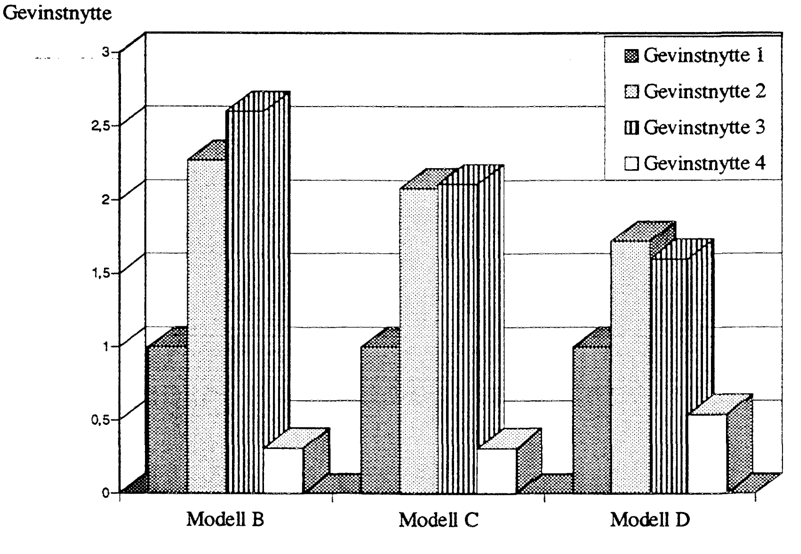
Tabell 5-1 : Resultater fra estimering av modell A-D

Modelltype	Modell A	Modell B	Modell C	Modell D
ln(L)	-1085.7	-1089.09	-1098.91	-1126.42
θ	-0.87(-2.48)	0.00(0)	1.00(∞)	10(∞)
$\hat{u}_1$	1.00(∞)	1.00(∞)	1.00(∞)	1.00(∞)
$\hat{u}_2$	2.27(6.53)	2.27(6.96)	2.08(7.54)	1.72(8.60)
$\hat{u}_3$	3.05(5.20)	2.60(6.50)	2.11(7.51)	1.60(7.96)
$\hat{u}_4$	0.45(5.77)	0.31(4.90)	0.31(3.08)	0.54(1.82)

I første rad er "loglikelihood"-verdien vist. Den er et mål på hvor sannsynlig vårt datamateriale er i modellens verden. Andre rad viser θ-verdiene i de ulike modellene. I modell A er som sagt, θ antatt å være en parameter i estimeringsprosessen. Oppgitt verdi er dermed kun et estimat. I rad fire til seks står parameterestimatene for nytten knyttet til de ulike gevinstene. Jeg satte apriori $\hat{u}_1=1$, noe som er vist i rad tre. I parentes står t-verdiene som beskriver forholdet mellom estimatorens forventningsverdi og standardavvik. Årsaken til at jeg har oppgitt t-verdiene i stedet for standardavvik, er at jeg vil ha et ordinalt mål. Med et ordinalt mål i denne sammenheng mener jeg et mål som kan sammenlignes uavhengig av verdien til estimatene.

Modell A er den mest generelle modellen. Parameterestimatet for θ er imidlertid negativt; nærmere bestemt er $\theta = -0.87$. Spørsmålet er om dette kan forsvares teoretisk. Becker krevde en strengt monotont økende funksjon h i sin modell, noe som er ensbetydende med en positiv θ-verdi. Negative θ kan derfor ikke forsvares på en teoretisk basis. Estimeringen må følgelig gjennomføres med den beskrankningen at $\theta \geq 0$. Som modell B-D indikerer, finnes det et optimum for $\theta = 0$ dersom den nye beskrankningen innføres. Det vil si at modell B er den av modellene som gir best tilpasning til hele observasjonsmaterialet. Nytteestimatene for modell B er vist i figur 5-3 sammen med estimatene for modell C og D.

Figur 5-3 : Paramerestimat i modell B, C og D



Problemet vi slet med i kapittel 4, var at det ikke var mulig å argumentere for om modell B var bedre enn modell C eller omvendt ved hjelp av teorien. Begge modellene spesifiserer en vurdering av et lotteri på grunnlag av forventet nytte. I avsnitt 4.2 beskrev jeg metoder for å sammenligne modeller. I det følgende vil jeg benytte meg av disse for å undersøke hvilken av modellene som gir best tilpasning til datamaterialet.

Jeg vil basere sammenligningen av modellene på tre mål; "likelihood"-verdien, "goodness of fit" og gjennomsnittlig prediksjonsfeil. Loglikelihoodsverdien er gitt ved

$$\ln\left(L\left(\theta,\hat{\mathbf{u}}\right)\right)=\sum_{n=1}^{1044}\left[\sum_{i=1}^{3}Y_{in}v_i\left(\theta,\hat{\mathbf{u}}\right)-\ln\left(\sum_{i=1}^{3}e^{v_i\left(\theta,\hat{\mathbf{u}}\right)}\right)\right] \tag{5.7}$$

der $\hat{\mathbf{u}}$ er en vektor bestående av estimatet for nytten knyttet til de fire utfallene. Det finnes to mål for "goodness of fit" (avsnitt 4.2). Jeg foretrekker her å bruke $\overline{\rho}^2$ siden antall parametre varierer.

Det er gitt ved

$$\bar{\rho}^2 = 1 - \frac{\ln\left(L\left(\theta, \hat{\mathbf{u}}\right)\right) - K}{\ln(L(\theta,0))} \tag{5.8}$$

hvor K er antall estimerte parametre. Jo høyere mål dess bedre. Et videre mål for hvor godt modellen beskriver observasjonsmaterialet, var gjennomsnittlig prediksjonsfeil (4.3). Den gjennomsnittlige prediksjonsfeilen skal være så liten som mulig. Alle mål er gitt i tabell 5-2.

Tabell 5-2: Sammenligningsmål

Modelltype	Modell A	Modell B	Modell C	Modell D
ln(L)	-1085.72	-1089.09	-1098.91	-1126.42
$\bar{\rho}^2$	0.0568	0.0478	0.0392	0.0153
$\bar{E}$	0.6254	0.6278	0.6382	0.6562

Jeg har argumentert for at θ bør være ikke-negativ for at modellen skal ha noen mening. Jeg vil derfor i det følgende se bort ifra modell A. Blant de resterende modellene viser resultatene som forventet, at modell B har den største "likelihood"- og "goodness of fit"-verdien og den minste gjennomsnittlige prediksjonsfeilen. Modell B ser ut til å være bedre enn modell C dersom vi kun betrakter disse verdiene. Spørsmålet er om dette er tilfeldig. Rangeringen er tilfeldig dersom den endres ved små endringer av datamaterialet. Modell B er tilnærmet lik en modell A dersom vi innfører beskrankningen $\theta \geq 0$. Dersom vi tillegg setter betingelsen at $\theta = 1$, blir resultatet modell C. LR-testen tester om en slik betingelse innebærer en signifikant dårligere modell. Nullhypotesen

$H_0 : \theta = 1$ mot $H_1 : \theta \neq 1$

ved hjelp av testobservatoren

$$LR = -2\left(\ln\left(L\left(0, \hat{\mathbf{u}}\right)\right) - \ln\left(L\left(1, \hat{\mathbf{u}}\right)\right)\right).$$

Observatoren er χ^2-fordelt med en frihetsgrad (en beskrankning). Vårt observasjonsmateriale gir LR=19.64, noe som gir grunn til forkastning på 0.05-nivå. På grunnlag av vårt datamateriale kan vi følgelig trekke to konklusjoner. For det første beskrives aktørenes valgadferd i dette eksperimentet best av en modell som er spesifisert ut fra antagelsen om at de maksimerer forventet nytte. Videre er denne antagelsen spesifisert bedre i modell B enn i modell C.

I modell B og C benytter aktøren i stor grad den informasjon som ligger i utfallsfordelingen. I modell D legger aktøren i større grad vekt på nytteverdiene i sin vurdering av et lotteri. I kapittel 3.3 sa jeg at en slik vurderingsmetode ikke var fornuftig i dette eksperimentet, i og med at gevinstene var like for alle de tre lotteriene. Vi bør kunne forvente at det gir seg utslag i at modell D får en dårligere forklaringsgrad. Vi ser at dette også er tilfelle når vi betrakter målene i tabell 5-2.

Til nå har vi kun estimert modeller på grunnlag av hele observasjonsmaterialet. Vi har antatt at alle aktørene kan beskrives av den samme modellen. I undersøkelsen deltok det ulike aktører. Det er rimelig å forvente at disse ni opplever forskjellig nytte knyttet til de ulike gevinstene. Da modellen ikke tar med attributter ved gevinstene eller karakteristika ved aktørene, er denne forskjellen i preferanser ikke implementert. Rimelig homogene preferanser er dermed en betingelse for at en modell estimert på hele observasjonsmaterialet, skal beskrive en aktør godt. Det er derfor essensielt å teste om preferansene er homogene. Dersom de ikke er det, bør vi estimere en modell for hver enkelt aktør.

5.4.3 Modeller for hver enkelt aktør

I forrige avsnitt antok vi at de ni aktørene som deltok i eksperimentet, hadde like preferanser og vurderingsmetode. Første antagelse forenkler modellen slik at vi bare får ett sett med nytteverdier. Disse er felles for alle aktører. Den andre antagelsen gir en felles θ. I dette avsnittet vil jeg undersøke om antagelsene holder.

5.4.3.1 Heterogene preferanser

Jeg vil i dette avsnittet undersøke om det er grunn til å anta at opplevd nytte av en gevinst varierer fra aktør til aktør. Det er det samme som å teste om en segmentering innebærer signifikant bedre modeller. En segmentering innebar at vi delte observasjonsmaterialet inn i flere undergrupper og estimerte en separat modell for hver undergruppe. Problemet i mange tilfeller, er å finne et godt segmenteringskriterie. I dette tilfellet er segmenteringskriteriet opplagt. Det er naturlig å dele observasjonsmaterialet inn i ni og estimere en modell for hver aktør.

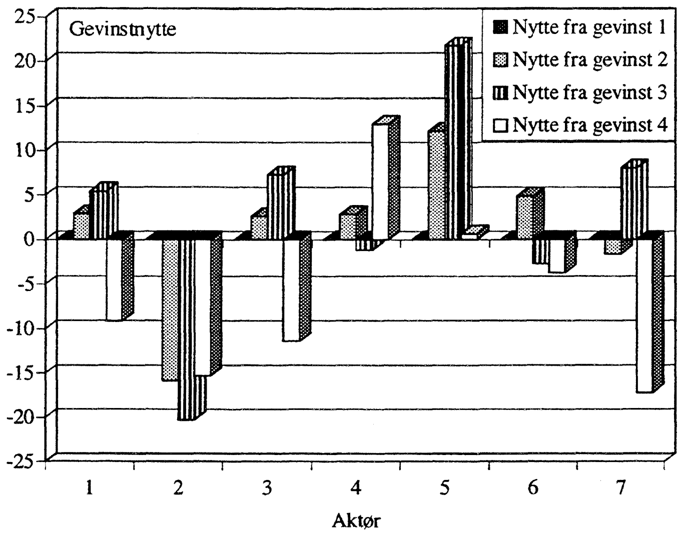
Jeg ser kun på om preferansene er heterogene i dette tilfellet. Jeg vil derfor fiksere θ for å la de individuelle nytteverdiene forklare forskjellige valg. Jeg vil bruke den θ som resulterte i den modellen med best forklaringsgrad for hele observasjonsmaterialet. Vi har ingen garanti for at denne verdien gir gode individuelle modeller, men den er det beste forslaget til en felles verdi. Det viste seg at det med θ=0 var mulig å estimere en individuell modell for syv av de ni aktørene. Årsaken til at det ikke var mulig å estimere en modell for de resterende to, vil jeg komme inn på i neste avsnitt. Resultater fra de syv estimeringene er gjengitt i tabell 5-3.

Tabell 5-3: Resultater fra estimering av syv individuelle modeller.

Aktør	1	2	3	4	5	6	7
ln(L)	-61.54	-30.35	-53.49	-53.81	-31.09	-67.48	-48.68
$\hat{u}_1$	0(∞)	0(∞)	0(∞)	0(∞)	0(∞)	0(∞)	0(∞)
$\hat{u}_2$	2.94(4.38)	-15.8(-3.97)	2.64(3.85)	2.85(3.99)	12.18(4.42)	4.89(6.26)	-1.61(-2.37)
$\hat{u}_3$	5.44(5.77)	-20.3(-4.32)	7.29(5.72)	-1.18 (1.18)	21.78(4.73)	-2.66(-3.43)	8.06(4.96)
$\hat{u}_4$	-9.11(-5.23)	-15.3(-3.71)	-11.4(-5.32)	13.0(5.90)	0.65(0.55)	-3.70(-3.62)	-17.2(-5.56)

Tabellen inneholder samme type informasjon som tabell 5-1; nemlig "loglikelihood"-verdi, estimater. T-verdier står i parentes. Figur 5-4 viser estimatene eller egentlig preferansene til de ulike aktørene.

Figur 5-4: Nytteestimater i modell B for syv aktører



Ved kun å se på figuren kan vi konstatere at preferansene ser svært ulike ut. Det er imidlertid også mulig å teste om de er signifikant forskjellig ved hjelp av en LR-test (avsnitt 4.2). Testen innebærer nullhypotesene

$$H_0 : \mathbf{u} = \mathbf{u}_t \qquad \text{mot} \qquad H_1 : \mathbf{u} \neq \mathbf{u}_t$$

der **u** er en vektor av parametre i modellen uten segmentering. For å finne disse parameterene er det nødvendig å estimere en felles modell for de syv aktørene det var mulig å estimere separate modeller for. Jeg estimerte den felles modellen for de 7 aktørene med θ som parameter. Denne modellen hadde ln(L)=831.59. $\mathbf{u}_t$ er parametre i en modell for segment $t \in \{1,2...,7\}$. De er estimat for de nytteverdiene aktør t tillegger gevinstene. Nullhypotesene testes simultant på grunnlag av testobservatoren

$$LR = -2\left(\ln\left(L\left(\theta,\hat{\mathbf{u}}\right)\right) - \sum_{t=1}^{7}\ln\left(L\left(0,\hat{\mathbf{u}}_t\right)\right)\right)$$

som er χ^2-fordelt med 17 frihetsgrader. Nullhypotesen forkastes følgelig på 0.05-nivå dersom LR>27.59. På grunnlag av vårt observasjonsmateriale blir LR = 968.3, noe som gir god grunn til forkastning. Det innebærer at referansene til aktørene er så ulike at de ikke kan beskrives av den samme modellen på en god måte. Individuelle modeller beskriver aktørene signifikant bedre.

Tabell 5-4: Sammenligningsmål for individuelle modeller

Aktør	1	2	3	4	5	6	7
lg(L)	-61.54	-30.35	-53.49	-53.81	-31.09	-67.48	-48.68
$\bar{\rho}^2$	0.4935	0.7383	0.5803	0.5577	0.7325	0.4469	0.5945
$\bar{E}$	0.3224	0.1513	0.2432	0.2389	0.1526	0.3248	0.2491

Aktørene beskrives i ulik grad av modellen. Dette sees av tabell 5-4. Aktør 2 og 5 er de som beskrives best. Disse to har en slående liten gjennomsnittlig prediksjonsfeil, kun drøye 0.15. Det vil si at modellen predikerer sannsynligheten for at de velger akkurat det de har valgt i observasjonsmaterialet, i gjennomsnitt til å være 0.85. Dersom vi betrakter "goodness of fit" for de ulike modellene, varierer de mellom 0.4469 og 0.7383. Til sammenligning hadde Ben-Akiva og Lerman (1985) "goodness of fit" - mål på maksimalt 0.45 i sin modell for valg av transportalternativ. Sett på bakgrunn av dette, må modellene sies å beskrive aktørene svært godt.

5.4.3.2 Ulik vurdering av et lotteri

Aktører som vurderer et lotteri forskjellig, må spesifiseres av separate modeller. Innenfor vår modellverden vil det si at den optimale θ varierer med aktøren. Becker et.al.(1963b) viste at dette var tilfelle for den valgsituasjonen de tok for seg. Det er derfor interessant å undersøke om det samme gjelder for aktørene som inngår i vårt observasjonsmateriale. Jeg forsøkte å estimere en modell tilsvarende modell A fra forrige avsnitt for hver enkelt aktør, for å se om optimal θ varierte. Estimeringen lyktes imidlertid ikke. Årsakene til dette kan være flere.

Rent konkret skal fem parametre estimeres på grunnlag av 116 observasjoner i en ulineær modell. Vi beveger oss med andre ord helt på grensen av det som er mulig.

Modellen bygger videre på en antagelse om at nytten er stokastisk for enten bare forskeren eller både forskeren og aktøren. Antagelsen er et forsøk på å forklare inkonsistens mellom det aktøren virkelig velger og det valg han ville ha gjort dersom han vurderte et lotteri på grunnlag av en transformasjon av deterministisk forventet nytte. Dersom en slik inkonsistens ikke er tilfelle, faller modellen på sin antagelse. Observasjonene inneholder ikke nok "irrasjonaltitet" eller variasjon til at en estimering er mulig.

En videre årsak til at estimeringen mislyktes kan være at strategien aktøren følger, ikke er inneholdt i modellen. Jeg spurte aktør 8 og 9 om hva de hadde vurdert lotteriene ut ifra. De ga følgende beskrivelse av sin vurderingsmetode. De rangerte de fire gevinstene. Aktør 8's rangering var eksempelvis

$$u_3 > u_2 > u_1 > u_4 .$$

I hver valgsituasjon valgte han det lotteri som ga høyest sannsynlighet for høyest prefererte gevinst, i dette tilfelle gevinst tre. Dersom den høyeste sannsynligheten var lik i to eller tre spill, valgte han det innenfor disse som ga høyest sannsynlighet for nest høyest prefererte gevinst og så videre. Det nærmeste vi kommer med å beskrive denne strategien med vår modell, er å tilordne nytteverdiene

$$u_1 = 0$$
$$u_2 = 0$$
$$u_3 = 1$$
$$u_4 = 0$$

til de fire gevinstene. un den høyest prefererte gevinsten gir et bidrag til nyttefunksjonen. Modell B beskriver vurderingsmetoden til aktør 8 så lenge han klarer å skille mellom lotteriene ved hjelp av den høyest prefererte gevinsten. Dersom utfallssannsynlighetene for denne er lik i to eller tre lotterier, vil nytten knyttet til disse også være lik. I vår modell er

$$P(V_j = V_i) = 0$$

per definisjon. Det vil si at lik nytte knyttet til et lotteri ikke forekommer. Vurderingsmetoden er følgelig ikke mulig å beskrive med modellen i slike tilfeller. En utvei er imidlertid å utelate alle spill der utfallssannsynligheten er lik for høyest prefererte gevinst. For denne aktøren var dette tilfelle i 22 av de 116 spillene. Jeg tok disse 22 spillene ut av observasjonsmaterialet og forsøkte å estimere modell A ut fra de resterende 94. Observasjonsmaterialet vil som en følge av dette ikke, inneholde opplysninger om videre rangering av gevinstene utover den første. Ikke uventet var det ikke nok variasjon i datamaterialet til å estimere en θ. Aktøren vurderer lotteriene konsekvent på grunnlag av forventet nytte. Han er løst sagt ikke inkonsistent nok til at vi får tilstrekkelig variasjon i data. Dette er et godt eksempel på at det er viktig å vurdere om de antagelsene en modell bygger på, virkelig holder i den situasjonen som betraktes.

Det er rimelig å anta at det som for de syv andre aktørene, også for denne aktøren er mulig å estimere en modell dersom θ fikseres til null, og dersom vi fortsatt utelater alle lotteriene der utfallssannsynligheten er lik for høyest prefererte gevinst. Resultatene er vist i tabell 5-5.

Tabell 5-5: Tilpasningsmål og estimater med t-verdier i parentes.

Aktør	8
ln(L)	-23.64
$\hat{u}_1$	3(∞)
$\hat{u}_2$	4.91(2.88)
$\hat{u}_3$	19.3(5.77)
$\hat{u}_4$	5.35(2.57)
$\bar{\rho}^2$	0.7711
$\bar{E}$	0.1175

"Goodness of fit" og gjennomsnittlig prediksjonsfeil indikerer en svært god modell.

En videre mulig grunn til at estimeringen mislykkes, er dersom aktøren har valgt helt tilfeldig. Jeg viste i avsnitt 3.3 hvordan dette kunne implementeres i modellen til Becker ved å la $\theta \to \infty$. Kort repetert velger aktøren det lotteriet som inneholder den gevinsten, det er knyttet høyest nytte til. I en situasjon som vår, der alle lotteriene inneholder de samme gevinstene, klarer ikke aktøren å skille mellom lotteriene på grunnlag av denne vurderingsmetoden. Hans valg blir helt tilfeldig. Han har ingen spesielle preferanser.

5.5 OPPSUMMERING

Jeg startet avsnitt 5.4 med å estimere en θ på grunnlag av hele observasjonsmaterialet og fant at θ ble estimert til null, gitt at parameteren apriori skulle ligge i intervallet $[0,\infty\rangle$. Resultatene er gitt i Tabell 5-1. Videre estimerte jeg modeller for θ lik henholdsvis 1 og 10. En LR-test viste at $\theta=1$ ga en signifikant dårligere modell enn $\theta=0$. En modell med $\theta = 10$ beskrev aktørenes valgadferd svært dårlig.

Forskjellige preferanser fanges ikke opp av modellen. Nok en LR-test viste at det var nødvendig å estimere en separat modell for hver enkelt aktør. Jeg klarte imidlertid ikke å estimere en modell for alle aktørene uten videre. For syv av aktørene var det mulig å estimere individuelle modeller innsatt $\theta=0$. Sammenligningsmålene i tabell 5-4 viser at de estimerte individuelle modellene beskriver aktørene i ulik grad. For de resterende to aktørene var det mulig å estimere en modell på grunnlag av et redusert datamateriale.

5.6 KONKLUSJON

Hvordan og på grunnlag av hvilke data en aktør vurderer et lotteri, kaller jeg for aktørens **vurderingsmetode**. En god modell må følgelig beskrive denne. Vurderingsmetoden kan variere med aktører og lotterityper. I kapittel 3 presenterte jeg med utgangspunkt i modellen til Becker et.al.(1963b) en modell (3.8) som var bestemt på en parameter nær. Ved å variere θ beskrev modellen ulike vurderingsmetoder. I avsnitt 3.3 beskrev jeg sammenhenger mellom vurderingsmetode og θ-verdi . En vanlig antagelse er at aktøren vurderer et lotteri utfra forventet nytte. Det finnes imidlertid flere mulige måter å beskrive denne vurderingsmetoden på , og det er ikke opplagt hvilke av dem som er best. Verdiene $\theta = 0$ og $\theta = 1$ resulterer begge i modellspesifikasjoner basert på forventet nytte. Siden det ikke er mulig å skille mellom disse spesifikasjonene på teoretisk basis, har jeg brukt datamateriale fra et eksperiment (vedlegg D og E) til å analysere hvilken av modellene som best gjenskaper observert valgadferd. Resultatene i avsnitt 5.4.2 viser at modell B beskriver observasjonsmaterialet bedre enn modell C. **Dersom vi forutsetter at aktøren vurderer et lotteri på grunnlag av en transformasjon av forventet nytte gitt ved (3.9) pluss et stokastisk restledd, ser det ut til at θ=0 gir det beste resultatet med datamaterialet som er brukt her.**

Det ikke mulig å si om dette resultatet gjelder generelt; uavhengig an aktør og rammebetingelser. Jeg vil i det følgende gå inn på årsaken til dette og hva som skal til for å generalisere konklusjonen.

Jeg klarte ikke å estimere individuelle modeller med θ som parameter. Siden vi ikke har noe sammenligningsgrunnlag, er det ikke mulig å si noe om modell B beskriver aktørenes valghandling best. Jeg nevnte en del årsaker til at det ikke var mulig å estimere modell A for hver enkelt aktør. For det første er 116 observasjoner ikke mye dersom fem parametre skal estimeres i en ulineær modell. Flere observasjoner kan gjøre en estimering mulig. Videre er det mulig at aktørene ikke valgte inkonsistent nok slik av datamaterialet ikke inneholdt nok variasjon. Løsningen på dette er å konstruere rammebetingelsene slik at aktørene får vanskeligheter med å evaluere dem konsistent for eksempel ved å øke mengden av informasjon aktøren må forholde seg til. Lotteriene kan eksempelvis bestå av forskjellige gevinster.

Eksperimentet innebærer en valgsituasjon med bestemte rammebetingelser; alle lotteriene ga mulighet til å vinne de samme gevinstene. Aktøren var derfor kun i stand til å skille lotteriene fra hverandre på grunnlag av utfallssannsynlighetene. $\theta \rightarrow \infty$ innsatt i modellen beskriver dermed en aktør som velger helt tilfeldig. I modell D legger aktøren i mindre grad vekt på den informasjonen som ligger i utfallssannsynlighetene. Dersom vi hadde konstruert et nytt eksperiment med like gevinstsannsynligheter og ulike gevinster, er det mulig at en tilsvarende estimering hadde vist at modell C eller D beskriver valgadferden best. På bakgrunn av dette er det følgelig ikke mulig å trekke en konklusjon som gjelder for alle rammebetingelser. Det er imidlertid mulig å undersøke om den gjelder mer generelt ved å gjøre andre eksperimenter.

REFERANSER

Becker, G.M., M.H. De Groot og J. Marschrak (1963a): "Stochastic Models and Choice Behavior". *Behavioral Sciences*, 8, s. 41-55.

Becker, G.M., M.H. De Groot og J. Marschrak (1963b): "An Experimenal Study and Some Stochastic Models for Wagers". *Behavioral Sciences*, 8, s. 199-202.

Ben-Akiva, M. og S.R. Lerman (1985): *"Discrete Choice Analysis"*. MIT Press, Cambridge.

Cramer, J.S (1991) : *"The Logit Model : An Introduction for Economists"*. Routhledge, Capman and Hall Inc., New York.

Cummins, C., H.B. Hall og R. Schnake (1991): *"TSP Reference Manual"*. Versjon 4.2. TSP International.

Dagsvik, J.K (1985) : " Kvalitativ Valghandlingsteori, en oversikt over feltet". *Sosialøkonomen*, 2, s. 32-38.

Dagsvik, J.K. (1995) : "Probabilistic Choice Models for Uncertain Outcomes". Discussion Paper nr.141. Statistisk Sentralbyrå.

Domencich, T og L. MacFadden (1975) : *"Urban Travel Demand - A Behavioral Analysis"*. North Holland, Amsterdam.

Edgell, S.E. og W.S.Geisler (1980) : "A Set-Theoretic Random Utility Model of Choice Behavior". *Journal of Mathematical Psychology*, 21, s.265-278.

Fechner, G.T. (1860) : *"Elemente der Psychophysik"*. Breitkopf und Hartel, Leibzig.

Hall, B.H. (1991): *"TSP User's Guide"*. Versjon 4.2. TSP International.

Hansen, A.C. (1995) : "Analyse av individers preferanser over lotterier basert på en stokastisk modell for sikre utfall". Sosialøkonomisk institutt, Universitet i Oslo.

Harless, D.W. og C.F. Camerer (1994) :"The Predictive Utility of Generalized Expected Utility Theories". *Econometrica*, 62, s 1251-1290.

Hey, J.D. og C. Orme (1994): " Investigating Generalizations of Expected Utility Theory using Experimental Data". *Econometrica*, 62, s 1291-1326.

Johnson, N.L. og S.Kotz (1970) : *"Continous Univariate Distributions"*, 1, s.272-289.Houghton Mifflin Company, Boston.

Luce, R.D. (1959) : *"Individual Choice Behavior: A Theoretical Analysis"*. Wiley, New York.

Luce, R.D. og P. Suppes (1965) : "Preferences, Utility and Subjective Probability". In R.D.Luce, R.R. Bush og E. Galanter (eds.), *Handbook of Mathematical Psychology*, III, s.249-410, Wiley, New York.

Manski, C. (1977) : "The Structure of Random Utility Models". *Theory and Decision*, 8, s 228-254.

McFadden, D. (1974) : "Conditional Logit Analysis of Qualitative Choice Behavior". In P. Zarembka (eds.), *Frontier in Econometrics*. Academic Press, New York.

Thurstone, L.L. (1927) : "A Law of Comparative Judgement". *Psychological Review*, 34, s. 272-286.

Thurstone, L.L. (1945) : "The Prediction of Choice". *Psychometrica*, 10, s.237-253.

Train, T. (1984) : *" Qualitative Choice Analysis: Theory, Economics and an Application to Automobile Demand"*. MIT Press, Cambridge.

VEDLEGG A: NOTASJONSOVERSIKT

MODELL FOR VALG MED SIKRE UTFALL

n : Indeks for aktør.
N : Totalt antall aktører.
i : Indeks for alternativ.
S : Totalt antall alternativer.
C_n : Valgmengden for aktør n, bestående av S alternativer.
U_{in} : Stokastisk nytte av alternativ i for aktør n.
u_{in} : Deterministisk nytte av alternativ i for aktør n.
ε_{in} : Nytterestledd, uobserverbare komponenter i nyttefunksjonen til aktør n for alternativ i.
$P_n(i;C_n)$: Sannsynligheten for at aktør n velger alternativ i fra mengden C_n.
$\boldsymbol{\beta}_n$: Parametervektor i den deterministiske nyttefunksjonen u_{in}, bestående av M elementer.
$\mathbf{x}_{in}$: Variabelvektor i den deterministiske nyttefunksjonen u_{in}, bestående av M elementer.
$\mathbf{s}_i$: Den del av variabelvektoren $\mathbf{x}_{in}$ som varierer med alternativene.
$\mathbf{w}_n$: Den del av variabelvektoren $\mathbf{x}_{in}$ som varierer med aktørene.
$\mathbf{h}(\mathbf{w}_n,\mathbf{s}_i)$: Den del av variabelvektoren $\mathbf{x}_{in}$ som varierer både med alternativene og aktørene.

MODELLER FOR LOTTERIER

Jeg har i utledningen av denne modellen utelatt aktørindeksen, n. Notasjonen bygger ellers på notasjonen for modeller for valg med sikre utfall.

i : Indeks for lotteri.
S : Totalt antall lotterier.
J : Stokastisk indeks for utfall eller gevinst.
C : Valgmengde, bestående av S lotterier.
K_i : Antall utfall eller gevinster i lotteri i.
D_i : Utfallsmengde, bestående av K_i gevinster.
g_{iJ} : Sannsynlighet for gevinst J i lotteri i.
U_J : Stokastisk nytte av gevinst J.
u_J : Deterministisk nytte av gevinst J.
η_{iJ} : Nytterestledd, uobserverbare komponenter i nyttefunksjonen for gevinst J i lotteri i.
V_i : Stokastisk nytte av lotteri i.
v_i : Deterministisk nytte av lotteri i.
ε_i : Nytterestledd, uobserverbare komponenter i nyttefunksjonen for lotteri i.

Generelt gjelder at:

- stokastiske variable beskrives med store bokstaver.
- deterministiske variable beskrives med små bokstaver.
- restledd variable beskrives med greske bokstaver.
- vektorer variable beskrives med fete symboler.

VEDLEGG B: EKSTREMVERDIFORDELINGENE

En stokastisk variabel gitt ved

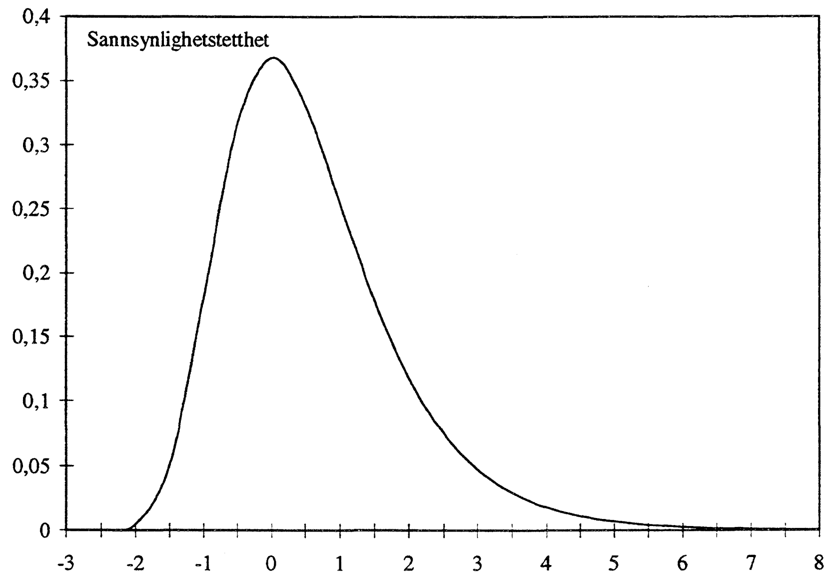
$$X = \underset{i}{\text{maks}}(U_i) \qquad \text{(C.1)}$$

er tilnærmet ekstremverdifordelt dersom i er stor og de stokastiske variablene U_i identisk og uavhengig fordelte. Det finnes tre typer av ekstremverdifordelinger, ofte kalt henholdsvis Gumbel-, Fréchet-, og Weibullfordelingen. Type I, Gumbelfordelingen, blir ofte refert til som ekstremverdifordelingen. En av grunnene til dette er at type II og III kan reduseres til type I ved enkle transformasjoner. På bakgrunn av dette og at type I er den som brukes i denne oppgaven, vil jeg kun gå inn på denne her. Dersom ε er Gumbelfordelt med parametrene (η,μ) er den kumulative sannsynlighetstettheten gitt ved

$$F(\varepsilon) = e^{-e^{-\mu(\varepsilon-\eta)}} \qquad \text{(C.2)}$$

der η er en lokalisering- og μ en skaleringsparameter. Sannsynlighetstettheten er vist i C-1.

Figur C-1: Gumbelfordelingen



Videre egenskaper ved fordelingen er

$$E(\varepsilon) = \eta + \frac{\gamma}{\mu}$$
$$\text{Var}(\varepsilon) = \frac{\pi^2}{6\mu^2}$$
$$E(\alpha\varepsilon + V) = \alpha E(\varepsilon) + V$$
$$\text{Var}(\alpha\varepsilon + V) = \alpha^2 \text{Var}(\varepsilon) \qquad \text{(C.3)}$$

der $\alpha>0$ og γ er Eulers konstant (~0.577).

Dersom $(\varepsilon_1, \varepsilon_2 \ldots\ldots \varepsilon_s)$ er uavhengige og gumbelfordelte med henholdsvis (η_1,μ), (η_2,μ)..... (η_s,μ), er

$$\underset{t\in\{1,s\}}{\text{maks}}(\varepsilon_t) \tag{C.4}$$

Gumbelfordelt med parametrene

$$\left(\frac{1}{\mu}\ln\left(\sum_{t=1}^{s} e^{\mu\eta_t}\right),\mu\right) \tag{C.5}$$

Vi ser at **Gumbelfordelingen er invariant over maksimeringsoperatoren**, noe som stemmer med (C.1). Denne egenskapen er svært nyttig i forbindelse med kvalitative valghandlingsmodeller.

Dersom ε_i og ε_j er Gumbelfordelt med henholdsvis (η_i,μ) og (η_j,μ), er ε_i-ε_j **logistisk** fordelt med

$$F(\varepsilon_i - \varepsilon_j) = \frac{1}{1 + e^{\mu(\eta_j - \eta_i - (\varepsilon_i - \varepsilon_j))}} \tag{C.6}$$

Dersom leseren er interessert i en mer inngående beskrivelse om ekstremverdifordelingene og den logistiske fordelingen, viser jeg til Johnson og Kotz (1970, kapittel 21 og 22).

VEDLEGG C: UTLEDNING AV MULTINOMISK LOGIT

Utledningen er tatt fra Domencich og MacFadden (1975). Logitmodeller bygger på en antagelse om at restleddene (ε_1, ε_2..........ε_s) er **uavhengig og identisk** gumbelfordelte med parametre (η,μ). Siden det kun er **forskjell** i nytteverdi som er av betydning, vil jeg som en forenkling anta at η=0. Sannsynligheten for at aktør n velger alternativ 1 er med utgangspunkt i (2.3) gitt ved

$$P_n(1;C_n) = P\left(u_{in} + \varepsilon_{in} \geq \underset{t\in\{2,s\}}{\text{maks}}(u_{tn} + \varepsilon_{tn})\right) = P\left(\underset{t\in\{2,s\}}{\text{maks}}(u_{tn} + \varepsilon_{tn}) - (u_{in} + \varepsilon_{in}) \leq 0\right) \tag{D.1}$$

der (u_{in}+ε_{in}) er Gumbelfordelt med parametrene (u_{in},μ), mens

$$\underset{t\in\{2,s\}}{\text{maks}}(u_{tn} + \varepsilon_{tn}) \tag{D.2}$$

ifølge C.5 er Gumbelfordelt med parametrene

$$\left(\frac{1}{\mu}\ln\left(\sum_{t=2}^{s} e^{\mu u_{tn}}\right), \mu\right) \tag{D.3}$$

Differansen mellom to Gumbelfordelte størrelser er logistisk fordelt. Ifølge C.6 er dermed

$$P_n(1;C_n) = P\left(\underset{t\in\{2,s\}}{\text{maks}}(u_{tn} + \varepsilon_{tn}) - (u_{1n} + \varepsilon_{1n}) \leq 0\right) = \frac{1}{1 + e^{\mu\left(\frac{1}{\mu}\ln\left(\sum_{t=2}^{s} e^{\mu u_{tn}}\right) - u_{1n}\right)}} = \frac{e^{\mu u_{1n}}}{\sum_{t=1}^{s} e^{\mu u_{tn}}}. \tag{D.4}$$

Det er ikke mulig å identifisere μ i sannsynlighetsfordelingen. Det derfor vanlig å velge μ=1 og i stedet skalere nyttefunksjonene for å få en optimal tilpasning til observasjonsmaterialet.

VEDLEGG D: PRESENTASJON AV EKSPERIMENTENE FOR AKTØRENE

Denne undersøkelsen går ut på å studere vår valgadferd når konsekvensene av de valg vi skal foreta, er ukjente på valgtidspunktet. Undersøkelsen består av 116 eksperimenter.

I hvert eksperiment er det tre spill; A, B og C.

I hvert av disse spillene er gevinsten en (1) reise.

Denne gevinsten blir trukket fra en mengde bestående av en eller flere av de fire reisene listet opp nedenfor:

- Reise 1 som går til Kairo (Egypt)
- Reise 2 som går til Beijing (Kina)
- Reise 3 som går til Bahamas (Karibien)
- Reise 4 som går til Dublin (Irland)

Hvilken av de fire mulige gevinstene som blir trukket ut, er usikkert. De respektive gevinstsannsynlighetene er imidlertid kjente.

Du kan anta at standarden på de fire reisene er den samme, for eksempel at alle reisene foregår på turistklasse, og at alle reisene inkluderer hotellopphold.

Jeg vil at du skal ta for deg ett og ett eksperiment. Forsøk å se for deg følgende situasjon: Gitt at kostnaden ved å delta i hvert av spillene A, B og C er den samme, for eksempel. 3200 kroner, hvilket av spillene A, B eller C ville du valgt å delta i ? Markèr ditt svar ved å avtegne en sirkel rundt det valgte alternativet (A, B eller C).

I denne undersøkelsen forutsetter vi at respondenten har et entydig førstevalg i hvert eksperiment. Dette medfører at når du er ferdig skal hvert eksperiment ha en og kun en markering. Husk også på at hvis du unnlater å svare på noen av eksperimentene, så kan ikke svarene dine brukes.

På forhånd takk !

116 EKSPERIMENTER

	Kairo	Beijing	Bahamas	Dublin
Eksperiment 1				
A	1	0	0	0
B	0.80	0.10	0.05	0.05
C	0.10	0.80	0.05	0.05
Eksperiment 2				
A	1	0	0	0
B	0.05	0.05	0.8	0.10
C	0.05	0.05	0.10	0.80
Eksperiment 3				
A	1	0	0	0
B	0.45	0.45	0.05	0.05
C	0.05	0.05	0.45	0.45
Eksperiment 4				
A	1	0	0	0
B	0.25	0.25	0.25	0.25
C	0.35	0.35	0.15	0.15
Eksperiment 5				
A	1	0	0	0
B	0.25	0.25	0.25	0.25
C	0.15	0.15	0.35	0.35
Eksperiment 6				
A	1	0	0	0
B	0.70	0.20	0.10	0
C	0.20	0.70	0.10	0
Eksperiment 7				
A	1	0	0	0
B	0.10	0.70	0.20	0
C	0.10	0.20	0.70	0
Eksperiment 8				
A	1	0	0	0
B	0	0.70	0.20	0.10
C	0	0.20	0.70	0.10
Eksperiment 9				
A	1	0	0	0
B	0	0.10	0.70	0.20
C	0	0.10	0.20	0.70
Eksperiment 10				
A	1	0	0	0
B	0.65	0.10	0.10	0.15
C	0.15	0.10	0.10	0.65
Eksperiment 11				
A	1	0	0	0
B	0.20	0.50	0.20	0.10
C	0.25	0.25	0.25	0.25
Eksperiment 12				
A	1	0	0	0
B	0.20	0.20	0.50	0.10
C	0.50	0.10	0.10	0.30
Eksperiment 13				
A	1	0	0	0
B	0.10	0.30	0.40	0.20
C	0.10	0.30	0.30	0.30
Eksperiment 14				
A	1	0	0	0
B	0.10	0.70	0.10	0.10
C	0.10	0.40	0.40	0.10
Eksperiment 15				
A	1	0	0	0
B	0.10	0.30	0.30	0.30
C	0.30	0.30	0.30	0.10
Eksperiment 16				
A	1	0	0	0
B	0.45	0.45	0.1	0
C	0	0.45	0.45	0.1
Eksperiment 17				
A	1	0	0	0
B	0.33	0.33	0.33	0.01
C	0.01	0.33	0.33	0.33

	Kairo	Beijing	Bahamas	Dublin
Eksperiment 18				
A	1	0	0	0
B	0.33	0.01	0.33	0.33
C	0.33	0.33	0.01	0.33
Eksperiment 19				
A	1	0	0	0
B	0.999	0.001	0	0
C	0.001	0.999	0	0
Eksperiment 20				
A	1	0	0	0
B	0	0	0.999	0.001
C	0	0	0.001	0.999
Eksperiment 21				
A	1	0	0	0
B	0	0.999	0.001	0
C	0	0.001	0.999	0
Eksperiment 22				
A	1	0	0	0
B	0.999	0	0	0.001
C	0.001	0	0	0.999
Eksperiment 23				
A	1	0	0	0
B	0.20	0.20	0.20	0.40
C	0.40	0.20	0.20	0.20
Eksperiment 24				
A	1	0	0	0
B	0.22	0.22	0.22	0.34
C	0.22	0.22	0.34	0.22
Eksperiment 25				
A	1	0	0	0
B	0.15	0.25	0.25	0.35
C	0.20	0.20	0.20	0.40
Eksperiment 26				
A	1	0	0	0
B	0.55	0.05	0.25	0.15
C	0.55	0.20	0.20	0.05
Eksperiment 27				
A	1	0	0	0
B	0.50	0.50	0	0
C	0	0	0.50	0.50
Eksperiment 28				
A	1	0	0	0
B	0.50	0	0	0.50
C	0	0.50	0.50	0
Eksperiment 29				
A	1	0	0	0
B	0.50	0	0.50	0
C	0	0.50	0	0.50
Eksperiment 30				
A	1	0	0	0
B	0.30	0.70	0	0
C	0	0	0.30	0.70
Eksperiment 31				
A	1	0	0	0
B	0.70	0.30	0	0
C	0	0	0.70	0.30
Eksperiment 32				
A	1	0	0	0
B	0	0.60	0.4	0
C	0	0.40	0.60	0
Eksperiment 33				
A	1	0	0	0
B	0	0.60	0.40	0
C	0.60	0	0.40	0
Eksperiment 34				
A	1	0	0	0
B	0	0.55	0.45	0
C	0.55	0	0.45	0
Eksperiment 35				
A	1	0	0	0
B	0.45	0.55	0	0
C	0.45	0	0.45	0.1

	Kairo	Beijing	Bahamas	Dublin
Eksperiment 36				
A	1	0	0	0
B	0	0.65	0	0.35
C	0.25	0.25	0.25	0.25
Eksperiment 37				
A	1	0	0	0
B	0	0.35	0.65	0
C	0.25	0.25	0.25	0.25
Eksperiment 38				
A	1	0	0	0
B	0	1	0	0
C	0	0	1	0
Eksperiment 39				
A	1	0	0	0
B	0	1	0	0
C	0	0	0	1
Eksperiment 40				
A	0.90	0.05	0.05	0
B	0.32	0.08	0.30	0.30
C	0.30	0.30	0.32	0.08
Eksperiment 41				
A	0.20	0.20	0.60	0
B	0.32	0.08	0.30	0.30
C	0.30	0.30	0.08	0.32
Eksperiment 42				
A	0.20	0.20	0.60	0
B	0.32	0.08	0.30	0.30
C	0.30	0.32	0.08	0.30
Eksperiment 43				
A	0.20	0.20	0.60	0
B	0.32	0.08	0.30	0.30
C	0.30	0.08	0.32	0.30
Eksperiment 44				
A	0.90	0	0.05	0.05
B	0.80	0.10	0.05	0.05
C	0.10	0.80	0.05	0.05
Eksperiment 45				
A	0.90	0	0.05	0.05
B	0.80	0.10	0.05	0.05
C	0.05	0.80	0.10	0.05
Eksperiment 46				
A	0.70	0.20	0	0.10
B	0.80	0.10	0.05	0.05
C	0.05	0.10	0.80	0.05
Eksperiment 47				
A	0.70	0.20	0	0.10
B	0.05	0.05	0.10	0.80
C	0.05	0.05	0.80	0.10
Eksperiment 48				
A	0.70	0.20	0	0.10
B	0.05	0.05	0.10	0.80
C	0.10	0.05	0.05	0.80
Eksperiment 49				
A	0.70	0.20	0	0.10
B	0.05	0.05	0.10	0.80
C	0.80	0.05	0.05	0.10
Eksperiment 50				
A	0.10	0.10	0.70	0.10
B	0.20	0.20	0.30	0.30
C	0.30	0.30	0.20	0.20
Eksperiment 51				
A	0.10	0.10	0.70	0.10
B	0.20	0.20	0.30	0.30
C	0.30	0.20	0.20	0.30
Eksperiment 52				
A	0.10	0.10	0.70	0.10
B	0.20	0.20	0.30	0.30
C	0.20	0.30	0.30	0.20
Eksperiment 53				
A	0.10	0.10	0.70	0.10
B	0.30	0.10	0.30	0.30
C	0.30	0.30	0.10	0.30

	Kairo	Beijing	Bahamas	Dublin
Eksperiment 54				
A	0.50	0.50	0	0
B	0.30	0.10	0.30	0.30
C	0.30	0.30	0.30	0.10
Eksperiment 55				
A	0.50	0.50	0	0
B	0.30	0.10	0.30	0.30
C	0.10	0.30	0.30	0.30
Eksperiment 56				
A	0	0.50	0.50	0
B	0.70	0.10	0.10	0.10
C	0.70	0.20	0.05	0.05
Eksperiment 57				
A	0	0.50	0.50	0
B	0.70	0.10	0.10	0.10
C	0.70	0.05	0.05	0.20
Eksperiment 58				
A	0	0	0.50	0.50
B	0.10	0.10	0.70	0.10
C	0.70	0.20	0.05	0.05
Eksperiment 59				
A	0	1	0	0
B	0.70	0.10	0.10	0.10
C	0.60	0.10	0.20	0.10
Eksperiment 60				
A	0	1	0	0
B	0.70	0.10	0.10	0.10
C	0.1	0.20	0.60	0.10
Eksperiment 61				
A	0	0.50	0.50	0
B	0.70	0.10	0.10	0.10
C	0.20	0.60	0.10	0.10
Eksperiment 62				
A	0	0.50	0.50	0
B	0.70	0.10	0.10	0.10
C	0.20	0.10	0.60	0.10
Eksperiment 63				
A	0	0.50	0.50	0
B	0.70	0.10	0.10	0.10
C	0.20	0.10	0.10	0.60
Eksperiment 64				
A	0	0	0.50	0.50
B	0.45	0.45	0.05	0.05
C	0.05	0.05	0.45	0.45
Eksperiment 6				
A	0	0	0.5	0.5
B	0.45	0.45	0.05	0.05
C	0.35	0.35	0.20	0.10
Eksperiment 66				
A	0	0	0.50	0.50
B	0. 05	0.05	0.45	0.45
C	0.65	0.05	0.20	0.10
Eksperiment 67				
A	0	0	0.50	0.50
B	0.05	0.45	0.45	0.05
C	0.10	0.2	0.65	0.05
Eksperiment 68				
A	0	0.50	0.50	0
B	0.45	0.05	0.05	0.45
C	0.10	0.20	0.65	0.05
Eksperiment 69				
A	0	0.50	0.50	0
B	0.40	0.10	0.10	0.40
C	0.10	0.20	0.65	0.05
Eksperiment 70				
A	0.4	0	0	0.60
B	0.10	0.40	0.10	0.40
C	0.10	0.20	0.65	0.05
Eksperiment 71				
A	0.40	0	0	0.60
B	0.10	0.40	0.40	0.10
C	0.10	0.20	0.65	0.05

	Kairo	Beijing	Bahamas	Dublin
Eksperiment 72				
A	0	0.40	0.60	0
B	0.50	0.20	0.20	0.10
C	0.10	0.20	0.65	0.05
Eksperiment 73				
A	0	0.40	0.60	0
B	0.10	0.20	0.20	0.50
C	0.10	0.20	0.65	0.05
Eksperiment 74				
A	0	0.40	0.20	0.40
B	0.20	0.50	0.10	0.20
C	0.10	0.20	0.65	0.05
Eksperiment 75				
A	0	0.40	0.20	0.40
B	0.20	0.10	0.50	0.20
C	0.10	0.20	0.65	0.05
Eksperiment 76				
A	0	0.40	0.20	0.40
B	0.15	0.15	0.50	0.20
C	0.15	0.15	0.20	0.50
Eksperiment 77				
A	0	0.40	0.30	0.30
B	0.50	0.20	0.15	0.15
C	0.20	0.50	0.15	0.15
Eksperiment 78				
A	0	0.40	0.30	0.30
B	0.15	0.50	0.20	0.15
C	0.15	0.20	0.50	0.15
Eksperiment 79				
A	0.90	0.10	0	0
B	0.50	0.15	0.15	0.20
C	0.20	0.15	0.15	0.50
Eksperiment 80				
A	0.90	0.10	0	0
B	0.25	0.25	0.25	0.25
C	0.20	0.20	0.40	0.20
Eksperiment 81				
A	0.90	0.10	0	0
B	0.25	0.25	0.25	0.25
C	0.40	0.2	0.20	0.20
Eksperment 82				
A	0	0	1	0
B	0.25	0.25	0.25	0.25
C	0.20	0.40	0.20	0.20
Eksperiment 83				
A	0	0	1	0
B	0.25	0.25	0.25	0.25
C	0.20	0.20	0.20	0.40
Eksperiment 84				
A	0	0.90	0.05	0.05
B	0.25	0.25	0.25	0.25
C	0.50	0.10	0.10	0.30
Eksperiment 85				
A	0	0.90	0.05	0.05
B	0.25	0.25	0.25	0.25
C	0.30	0.10	0.10	0.50
Eksperiment 86				
A	0	0.90	0.05	0.05
B	0.25	0.25	0.25	0.25
C	0.50	0.10	0.30	0.10
Eksperiment 87				
A	0.20	0.20	0.20	0.40
B	0.25	0.25	0.25	0.25
C	0.10	0.50	0.10	0.30
Eksperiment 88				
A	0.20	0.80	0	0
B	0.25	0.1	0.35	0.30
C	0.30	0.35	0.10	0.25
Eksperiment 89				
A	0.20	0.80	0	0
B	0.25	0.10	0.35	0.30
C	0.25	0.35	0.10	0.30

	Kairo	Beijing	Bahamas	Dublin
Eksperiment 90				
A	0.20	0.80	0	0
B	0.20	0.25	0.25	0.30
C	0.30	0.25	0.25	0.20
Eksperiment 91				
A	0	0.85	0	0.15
B	0.90	0	0	0.10
C	0.90	0.10	0	0
Eksperiment 92				
A	0.90	0	0.10	0
B	0.80	0	0.20	0
C	0.80	0.20	0	0
Eksperiment 93				
A	0.90	0	0.10	0
B	0.50	0.30	0.20	0
C	0.5	0.10	0.30	0.10
Eksperiment 94				
A	0.85	0	0	0.15
B	0.70	0.10	0.10	0.10
C	0.10	0.60	0.20	0.10
Eksperiment 95				
A	0.15	0	0	0.85
B	0.60	0.10	0.20	0.10
C	0.60	0.20	0.10	0.10
Eksperiment 96				
A	0.15	0	0	0.85
B	0.50	0.20	0.30	0
C	0.50	0.30	0.20	0
Eksperiment 97				
A	0	0.85	0	0.15
B	0.20	0.10	0.70	0
C	0.10	0.10	0.70	0.10
Eksperiment 98				
A	0	0.65	0	0.35
B	0.05	0.05	0.80	0.10
C	0.10	0.10	0.70	0.10
Eksperiment 99				
A	0	0.65	0	0.35
B	0.30	0.30	0.30	0.10
C	0.40	0.40	0.10	0.10
Eksperiment 100				
A	0	0.35	0	0.65
B	0.20	0.20	0.50	0.10
C	0.60	0.40	0	0
Eksperiment 101				
A	0	0.45	0	0.55
B	0.10	0.65	0.25	0
C	0.15	0.15	0.70	0
Eksperiment 102				
A	0	0.35	0	0.65
B	0.10	0.65	0.25	0
C	0.70	0.15	0.15	0
Eksperiment 103				
A	0	0.35	0	0.65
B	0.13	0.17	0.67	0.03
C	0.03	0.77	0.10	0.10
Eksperiment 104				
A	0	0.45	0	0.55
B	0.13	0.13	0.13	0.61
C	0.61	0.13	0.13	0.13
Eksperiment 105				
A	0	0.45	0.55	0
B	0.20	0.20	0.30	0.30
C	0.25	0.25	0.25	0.25
Eksperiment 106				
A	0	0.75	0	0.25
B	0.30	0.30	0.20	0.20
C	0.25	0.25	0.25	0.25
Eksperiment 107				
A	0	0.75	0	0.25
B	0.25	0.25	0.25	0.25
C	0.05	0.05	0.80	0.10

	Kairo	Beijing	Bahamas	Dublin
Eksperiment 108				
A	0	0	0.75	0.25
B	0.25	0.25	0.25	0.25
C	0.05	0.80	0.05	0.10
Eksperiment 109				
A	0	0	0.75	0.25
B	0.25	0.25	0.25	0.25
C	0.80	0.05	0.05	0.10
Eksperiment 110				
A	0	0	0.25	0.75
B	0	0.50	0.50	0
C	0	1	0	0
Eksperiment 111				
A	0	0	0.25	0.75
B	0	0.50	0.50	0
C	0	0	1	0
Eksperiment 112				
A	0.25	0.75	0	0
B	0	1	0	0
C	0.50	0	0	0.50
Eksperiment 113				
A	0.25	0.75	0	0
B	0.50	0.50	0	0
C	0	0	0.50	0.50
Eksperiment 114				
A	0	0	1	0
B	0.50	0.50	0	0
C	0.50	0	0.50	0
Eksperiment 115				
A	0	1	0	0
B	0.50	0.50	0	0
C	0	0.50	0	0.50
Eksperiment 116				
A	1	0	0	0
B	0	0	0	1
C	0	1	0	0

VEDLEGG E: OVERSIKT OVER DATAMATERIALET

Aktørenes svar er gjengitt i tabellene nedenfor.

Aktør **Eksperiment**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	3	1	2	3	3	3	2	3	1
2	2	1	2	3	2	2	2	2	1
3	2	1	2	3	3	2	2	3	1
4	3	1	3	2	2	3	3	2	1
5	2	1	1	3	3	2	2	3	1
6	3	1	3	3	3	3	2	3	1
7	3	1	3	2	3	2	3	3	1
8	3	1	3	2	3	2	3	3	1
9	2	1	2	3	2	2	2	2	1
10	2	1	1	3	2	2	2	2	1
11	2	1	2	3	2	2	2	3	1
12	2	1	2	3	2	2	2	2	1
13	2	1	2	3	2	2	1	2	1
14	3	1	3	2	3	2	3	3	1
15	3	1	3	2	3	3	3	3	1
16	3	1	3	3	3	2	3	3	1
17	2	1	2	3	2	2	2	2	1
18	2	1	1	3	2	3	1	2	1
19	3	1	3	3	3	3	2	3	3
20	2	1	2	3	2	1	2	2	1
21	3	1	3	2	3	2	3	3	1
22	2	1	1	3	2	1	1	1	3
23	1	1	3	2	3	3	1	3	1
24	3	1	1	2	3	1	1	3	1
25	2	1	1	3	2	1	1	2	1
26	2	1	3	2	3	3	3	2	1
27	2	1	2	3	3	2	2	3	2
28	3	1	3	2	3	3	3	3	2
29	2	1	2	3	2	2	2	2	3
30	2	1	2	3	2	2	1	3	2
31	2	1	2	2	3	2	1	3	2
32	3	1	3	2	3	2	3	3	1
33	2	1	2	2	2	2	3	2	1
34	2	1	2	2	2	2	3	2	1
35	2	1	2	2	3	2	3	3	2
36	3	1	1	2	3	2	1	3	2
37	2	1	2	3	2	3	2	2	1
38	3	1	3	2	3	2	3	3	2
39	2	1	2	3	2	2	1	2	3
40	3	1	3	2	3	1	3	3	1
41	1	3	1	3	1	1	1	1	3
42	1	3	1	3	1	3	1	1	3
43	1	2	1	2	1	1	1	1	2
44	3	1	3	3	3	3	1	3	3
45	3	1	3	3	3	3	1	3	2
46	3	2	3	1	3	1	3	3	1
47	3	1	3	2	3	1	3	3	1
48	1	1	2	3	2	1	1	2	1
49	1	3	2	2	2	1	3	2	1
50	1	3	1	2	1	3	1	1	3
51	1	3	1	3	1	3	1	1	3
52	1	3	1	2	1	3	1	1	2
53	1	3	1	3	1	3	1	1	3
54	1	3	3	2	3	1	3	3	1
55	1	1	1	3	3	1	2	3	1
56	1	3	1	2	1	3	1	1	3
57	1	2	1	3	1	2	1	1	3

Aktør Eksperiment	1	2	3	4	5	6	7	8	9
58	2	3	2	1	2	3	2	2	3
59	3	2	3	2	1	2	3	3	1
60	3	2	1	3	3	2	3	3	1
61	1	2	1	3	1	3	1	1	3
62	1	2	1	2	1	2	3	3	2
63	1	2	1	3	1	2	1	1	3
64	3	2	2	1	1	2	2	1	2
65	3	2	3	1	1	2	3	1	2
66	3	3	3	1	1	3	3	1	3
67	2	2	3	1	3	2	3	3	2
68	3	2	3	2	1	1	3	3	2
69	3	2	3	2	1	1	3	3	2
70	3	1	3	1	3	2	3	3	1
71	3	1	3	1	3	2	3	3	1
72	1	2	3	2	1	2	3	3	2
73	1	2	3	2	1	3	3	3	2
74	3	2	3	1	3	2	3	3	2
75	3	2	3	1	3	1	3	3	1
76	2	3	2	3	2	1	2	2	1
77	2	2	3	1	1	2	2	1	3
78	3	2	3	1	3	2	3	3	2
79	2	1	1	3	3	2	1	2	1
80	1	1	1	2	3	1	1	3	1
81	1	1	1	2	2	1	1	2	1
82	1	3	1	2	1	3	1	1	3
83	1	2	1	3	1	2	1	1	3
84	1	3	1	3	1	1	1	2	1
85	1	3	1	3	1	1	1	2	1
86	1	3	1	2	1	1	3	3	1
87	3	3	2	1	3	3	2	2	3
88	1	3	1	2	2	1	1	2	1
89	1	3	1	3	2	1	1	2	1
90	1	3	1	2	2	1	1	3	1
91	3	2	3	1	1	1	3	1	1
92	2	1	2	3	2	3	2	2	3
93	2	1	2	3	3	2	3	3	1
94	2	1	3	1	3	3	2	3	1
95	2	3	2	1	2	3	2	2	1
96	2	3	2	1	2	3	2	2	1
97	2	3	2	1	3	1	2	2	1
98	2	1	2	1	2	1	2	2	1
99	2	3	2	1	2	1	2	2	1
100	3	3	2	1	2	3	2	2	1
101	2	1	3	1	3	1	3	3	1
102	2	3	2	1	2	2	3	2	1
103	2	1	2	1	2	3	2	2	1
104	3	3	3	2	1	3	3	3	1
105	1	3	1	2	1	1	1	1	3
106	2	2	3	1	1	2	2	3	1
107	3	2	3	1	3	1	3	3	1
108	3	2	1	2	1	3	3	1	3
109	3	3	1	2	1	3	3	1	3
110	2	2	2	1	3	3	2	2	3
111	2	2	3	1	2	2	3	3	1
112	1	3	2	3	2	2	1	2	2
113	2	2	1	3	3	1	2	3	1
114	3	2	1	1	1	2	1	1	2
115	2	2	1	2	1	1	2	1	3
116	3	1	3	3	3	3	1	3	2

VEDLEGG F: PROGRAMKODE

Programkoden er kun et eksempel. Den estimerer modell A med beskrankningen $\theta \geq 0$ på grunnlag av hele observasjonsmaterialet. Resultatene av en slik kjøring er å finne i vedlegg F. Deler av koden må endres dersom jeg ønsker å estimere en annen modell eller samme modell på grunnlag av andre data. Jeg har uthevet de delene av koden som må forandres. Tegnet ? angir kommentar. Teksten bak dette tegnet leses ikke av TSP.

```
OPTIONS CRT,limcol=90;
TITLE 'ESTIMERING AV ALLE AKTØRER MED θ SOM PARAMETER';
FREQ N;

? Angir størrelsen av datamaterialet, det vil si antall observasjoner
set n=1044;
smpl 1,n;

? Leser inn nummer på valgte lotteri i serien v
LOAD(FILE='./valg9.df') nr v;

? Leser datafil med gevinstsannsynligheter inn i dataserier
LOAD(FILE='./spill9.df')gA1 gA2 gA3 gA4 gB1 gB2 gB3 gB4 gC1 gC2 gC3 gC4;

? Genererer responsvariable
GENR Y1=(V=1);
GENR Y2=(V=2);
GENR Y3=(V=3);

? Sett konstanter
b1=1;
? parameter=0  gir modell C

? Beskranker t til å være positiv. (t tilsvarer θ i modellen.)
FRML expt t=exp(parameter);

? Definerer nyttefunksjonene
FRML expt1 V1=(1/t)*log(b1**t*gA1+b2**t*gA2+b3**t*gA3+b4**t*gA4);
FRML expt2 V2=(1/t)*log(b1**t*gB1+b2**t*gB2+b3**t*gB3+b4**t*gB4);
FRML expt3 V3=(1/t)*log(b1**t*gC1+b2**t*gC2+b3**t*gC3+b4**t*gC4);

? Substituerer beskrankningen inn i nyttefunksjonene ved hjelp av en løkke
dot 1,2,3;
EQSUB expt. expt;
enddot;

? Definerer "likelihood"-funksjonen
FRML llike loglike=Y1*V1+Y2*V2+Y3*V3-log(exp(V1)+exp(V2)+exp(V3));

? Substituerer nyttefunksjonene inn i "likelihoodfunksjonen"
EQSUB llike expt1 expt2 expt3;

? Leser inn parametrene som skal estimeres og deres start-verdier
PARAM parameter,0.5 b2,2 b4,0.3 b3,2;

? Estimerer parametrene i modellen etter sannsynlighetsmaksimeringsprinsippet.
ML(HCOV=N,HITER=N,maxit=200, tol=.0001) llike;

? Skriver ut kovariansmatrisen til estimatene
print @vcov;

? Beregner nytte utfra estimerte parametre og lagrer svar i v1,v2,v3
GENR expt1;
GENR expt2;
GENR expt3;

? Definerer formlene for valgsannsynlighetene
FRML P1 pa=exp(V1)/(exp(V1)+exp(V2)+exp(V3));
FRML P2 pb=exp(V2)/(exp(V1)+exp(V2)+exp(V3));
FRML P3 pc=exp(V3)/(exp(V1)+exp(V2)+exp(V3));
```

```
? Definerer formel for prediksjonsavvik og substituerer inn valgsannsynlighetene
FRML Res e=1-((v=1)*pa+(v=2)*pb+(v=3)*pc);
EQSUB Res P1 P2 P3;

? Beregner prediksjonsavvik og valgsannsynligheter
GENR Res;
GENR P1;
GENR P2;
GENR P3;

? Finner gjennomsnittlig prediksjonsavvik og skriver det ut
MSD e;
```

? Beregner $\overline{\rho}^2$ og ρ^2 ved først å beregne "likelihood"-verdien for så å skrive dem ut.

```
GENR llike;
MSD loglike;
SET rhosqu=1+((@sum-4)/(n*log(3)));
SET rho=1+((@sum)/(n*log(3)));
print rho, rhosqu ;
```

VEDLEGG G: EKSEMPEL PÅ UTSKRIFT AV EN ESTIMERING

Nedenfor følger resulatet av kjøring av programkoden i vedledd F. Modell A med beskrankning estimeres på grunnlag av hele observasjonsmaterialet.

```
          ESTIMERING AV ALLE AKTØRER  MED T SOM PARAMETER
          ***********************************************
Current sample:  1 to 1044

          MAXIMUM LIKELIHOOD ESTIMATION
          *****************************
EQUATION: LLIKE

Working space used: 37435
               STARTING VALUES

          TETTA      B2        B3        B4
VALUE    -0.50000   2.00000   2.00000   0.50000

IN OBSERVATION    1 COMPUTING LOG LIKELIHOOD.

F=  1097.6   FNEW=  1092.7   ISQZ=  1 STEP= 0.50000   CRIT=  21.452
F=  1092.7   FNEW=  1089.6   ISQZ=  0 STEP=  1.0000   CRIT=  5.6304
F=  1089.6   FNEW=  1089.3   ISQZ=  0 STEP=  1.0000   CRIT= 0.54070
F=  1089.3   FNEW=  1089.2   ISQZ=  0 STEP=  1.0000   CRIT= 0.16341
F=  1089.2   FNEW=  1089.1   ISQZ=  0 STEP=  1.0000   CRIT= 0.61695E-01
F=  1089.1   FNEW=  1089.1   ISQZ=  0 STEP=  1.0000   CRIT= 0.22721E-01
F=  1089.1   FNEW=  1089.1   ISQZ=  0 STEP=  1.0000   CRIT= 0.77909E-02
F=  1089.1   FNEW=  1089.1   ISQZ=  0 STEP=  1.0000   CRIT= 0.17404E-02
F=  1089.1   FNEW=  1089.1   ISQZ=  0 STEP=  1.0000   CRIT= 0.50418E-04
F=  1089.1   FNEW=  1089.1   ISQZ=  0 STEP=  1.0000   CRIT= 0.14696E-08

CONVERGENCE ACHIEVED AFTER  10 ITERATIONS

 21 FUNCTION EVALUATIONS.
        LOG OF LIKELIHOOD FUNCTION =  -1089.09
          NUMBER OF OBSERVATIONS =  1044
```

Parameter	Estimate	Standard Error	t-statistic
TETTA	**-8.51668**	**18.0024**	**-.473087**
B2	**2.27127**	**.326170**	**6.96346**
B3	**2.60364**	**.400369**	**6.50308**
B4	**.340484**	**.069423**	**4.90447**

```
Standard Errors computed from  analytic second derivatives
(Newton)

               @VCOV

        1              2              3              4
  1  324.08501
  2  0.0093962      0.10639
  3  -0.015461      0.037284      0.16030
  4  0.0030899      0.0043983     0.0036914      0.0048196


          RESULTS OF COVARIANCE PROCEDURE
          ******************************

NUMBER OF OBSERVATIONS: 1044

          MEAN      STD DEV    MINIMUM    MAXIMUM
E        0.62780    0.11342    0.34043    0.91339


          SUM     VARIANCE
```

E	655.42272	0.012863

RESULTS OF COVARIANCE PROCEDURE
****

NUMBER OF OBSERVATIONS: 1044

	MEAN	STD DEV	MINIMUM	MAXIMUM
LOGLIKE	-1.04319	0.35046	-2.44640	-0.41617

	SUM	VARIANCE
LOGLIKE	**-1089.09340**	0.12282

RHO = 0.050445
RHOSQU = **0.046957**

Utkommet i serien Notater fra Forskningsavdelingen

94/11 E. Holmøy og B. Strøm: Virkningsberegninger på MGS-5, 1991-versjonen

94/12 K.Ø. Sørensen: En databank med fylkesfordelte nasjonalregnskapstall

94/13 B. Holtsmark: Tjenesteytende virksomhet i Norge. Revidert versjon, august 1994

94/15 T. Eika, S.I. Hove og L. Haakonsen: KVARTS i praksis. Macro-systemer og rutiner

94/17 E. Bowitz og I. Holm: Nye relasjoner i MODAG, januar 1994. Teknisk dokumentasjon

94/18 Y. Vogt: Innføring i FAME

94/22 M.W. Arneberg: LOTTE-TRYGD. Teknisk dokumentasjon

95/5 D. Fredriksen: MOSART Teknisk dokumentasjon

95/7 K. Olsen: Nytte- og kostnadsvirkninger av en norsk oppfyllelse av nasjonale utslippsmålsettinger

95/15 T. Karlsen: Opptimal karbonbeskatning og virkningen på norsk petroleumsformue

95/17 Å. Cappelen, T. Skjerpen og J. Aasness: Konsumetterspørsel, tjenesteproduksjon og sysselsetting. En mikro til makroanalyse

95/24 H.T. Mysen: Nordisk energimarkedsmodell. Dokumentasjon av delmodell for energietterspørsel i industrien

95/26 I. Aslaksen, T. Fagerli og H.A. Gravningsmyhr: Produksjon og konsum i husholdningene

95/29 B.E. Naug: Eksport- og importlikninger i KVARTS

95/31 B.E. Naug: Etterspørsel etter arbeidskraft - en litteraturoversikt

95/35 T.J. Klette: Vekst og produktivitet i norsk industri. Hovedrapport fra et NFR-prosjekt

95/40 L. Lerskau: Oversikt over konjunkturindikatorer i databasen NORMAP og FAME

95/46 B.E. Naug: Estimering av eksportrelasjoner på disaggregerte kvartalsdata

95/47 K. Moum: Beregning av bruttoproduksjon og eierinntekt i boligsektoren i nasjonalregnskapet - noen metodiske synspunkter

95/52 T. Kornstad: Simulering av konsum og arbeidstilbud i et livsløpsperspektiv

95/56 A. Langørgen: Faktorer bak kommunale variasjoner i utgifter til sosialhjelp og barnevern

95/58 T. W. Karlsen: Energimarkedet fra 1973 og fram mot 2010

96/3 I. M. Smestad: Valg under usikkerhet: En analyse av eksperimentdata basert på kvalitative valghandlingsmodeller

Statistisk sentralbyrå

Oslo
Postboks 8131 Dep.
0033 Oslo

Telefon: 22 86 45 00
Telefaks: 22 86 49 73

Kongsvinger
Postboks 1260
2201 Kongsvinger

Telefon: 62 88 50 00
Telefaks. 62 88 50 30

ISSN 0806-3745

Statistisk sentralbyrå
Statistics Norway