

RAPPORTER

80/26

**LINEÆRE OG LOG-LINEÆRE MODELLER
FOR KVALITATIVE AVHENGIGE
VARIABLE**

LINEAR AND LOG-LINEAR QUALITATIVE RESPONSE MODELS

AV/BY
LASSE FRIDSTRØM

**STATISTISK SENTRALBYRÅ
OSLO**

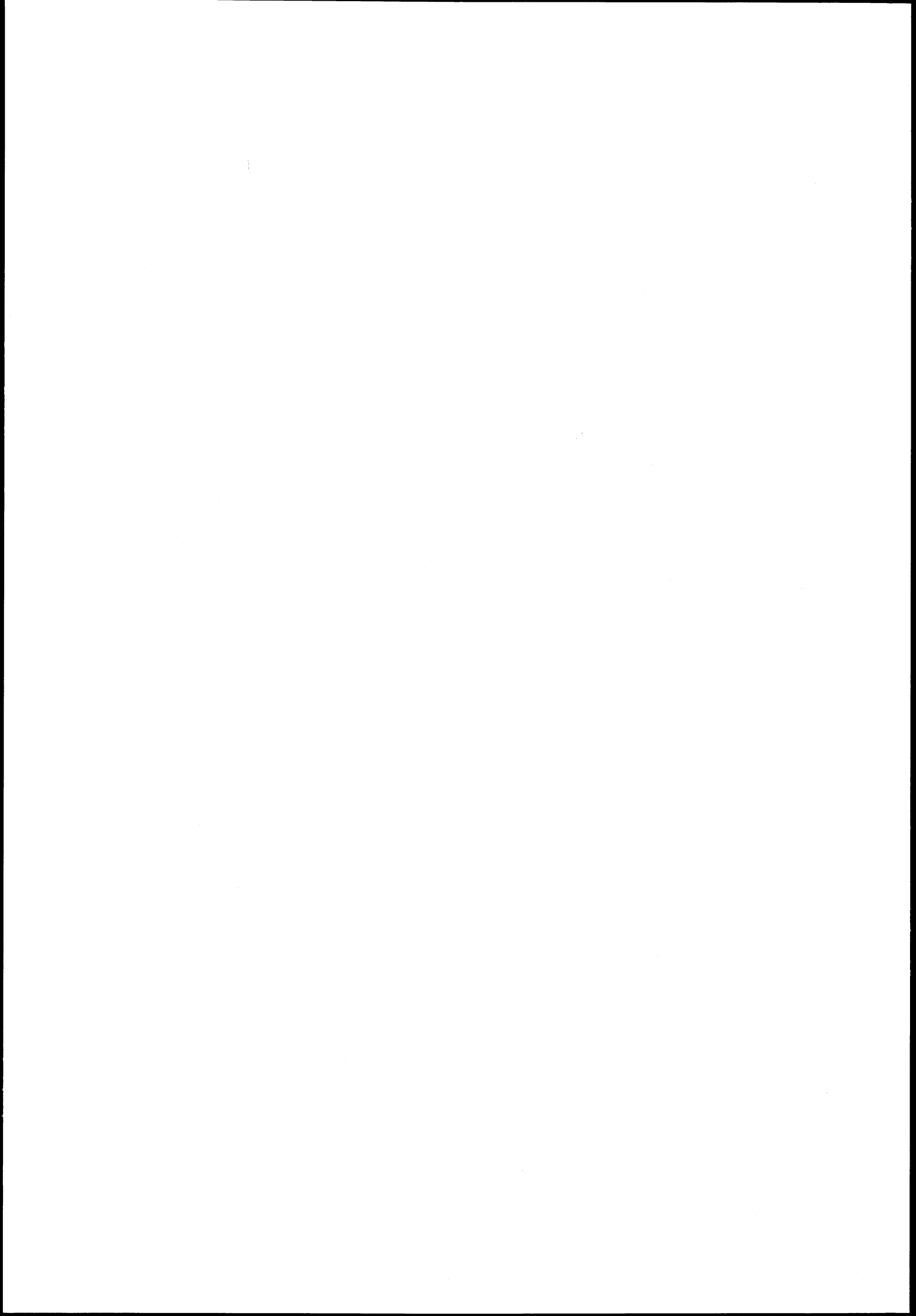
RAPPORTER FRA STATISTISK SENTRALBYRÅ 80/26

LINEÆRE OG LOG-LINEÆRE MODELLER
FOR KVALITATIVE AVHENGIGE VARIABLE

LINEAR AND LOG-LINEAR
QUALITATIVE RESPONSE MODELS

AV/BY
LASSE FRIDSTRØM

OSLO 1980
ISBN 82-537-1184-0
ISSN 0332-8422



FORORD

Denne rapporten gir et oversyn over en del statistiske analysemetoder som kan nyttes når en kvalitativ variabel blir søkt forklart ved hjelp av en rekke andre kvalitative og/eller kvantitative kjennetegn. En del av analysemetodene blir nå utprøvd på et datamateriale fra Arbeidskraftundersøkelsene 1976.

Statistisk Sentralbyrå, Oslo, 3. september 1980

Petter Jakob Bjerve

Per Sevaldson

PREFACE

This publication is a survey of qualitative response models and related statistical inference methods. A data set from the Norwegian Labour Force Sample Surveys 1976 is used to contrast some of the techniques empirically.

Central Bureau of Statistics, Oslo, 3 September 1980

Petter Jakob Bjerve

Per Sevaldson

INNHold

	Side
1. Innledning og disposisjon	13
2. Bakgrunn og formål	14
3. Modeller for kvalitative avhengige variable	16
4. Den lineære sannsynlighetsmodell	20
4.1. Binær avhengig variabel	20
4.2. Multinær avhengig vektor	21
4.3. Estimering ved to-trinns generaliserte minste kvadraters metode (2SGLS)	25
4.4. Sammenhengen mellom individdata og grupperte data	28
4.5. Estimering ved vanlig minste kvadraters metode (OLS)	29
4.6. Estimering ved sannsynlighetsmaksimering (ML)	31
4.7. Estimering ved minimum Neyman kji-kvadrat ($\min \chi_N^2$)	32
4.8. Sammenlikning mellom OLS og 2SGLS	35
5. Den log-lineære sannsynlighetsmodell	35
5.1. Binær avhengig variabel	35
5.2. Multinær avhengig vektor	38
5.3. Estimering ved to-trinns generaliserte minste kvadraters metode (2SGLS)	39
5.4. Estimering ved sannsynlighetsmaksimering (ML)	42
5.4.1. Multinomisk logit-modell	42
5.4.2. Symmetrisk log-lineær formulering	44
5.5. Hierarkiske hypoteser for den log-lineære modell	44
5.6. Estimering ved minimum logit kji-kvadrat ($\min \chi_L^2$)	47
5.7. Sammenlikning mellom 2SGLS og ML	48
6. Probit-modellen	50
7. Føyningsmål for lineære og log-lineære sannsynlighetsmodeller	53
7.1. Determinasjonskoeffisienten (R^2) for den lineære modell	53
7.1.1. Generelt om OLS-estimering	53
7.1.2. Spesielt om binær regresjon	54
7.1.2.1. Teoretisk drøfting	54
7.1.2.2. Regneeksempler	61
7.2. Kji-kvadrat-mål	65
7.2.1. Pearson kji-kvadrat (χ_P^2)	65
7.2.2. Neyman kji-kvadrat (χ_N^2)	66
7.2.3. Kji-kvadrat-observatoren basert på sannsynlighetskvoten (G^2)	66
7.3. Feilklassifiseringsindeksen (C)	66
7.4. Sammenfatning om føyningsmål	67
8. Opplegg for den empiriske analysen	67
8.1. Omfang	67
8.2. Datamaterialet	67
8.3. Problemet med tomme celler	70
8.4. Programutrustning	71

INNHold (forts.)

	Side
9. Beregningsresultater	71
9.1. Sammenlikning av føyningsmål	71
9.2. Variansen til parameterestimatorene i den lineære modell. Sammenlikning mellom OLS og 2SGLS	74
9.3. Hva tjener vi på å bruke en sannsynlighetsmodell?	76
9.4. Hypoteseprøving	78
9.4.1. Log-lineær modell	78
9.4.2. Lineær modell	80
9.4.3. Sammenlikning mellom lineær og log-lineær modell	81
10. Forsøk på en syntese	81
10.1. A priori-argumenter	81
10.2. Praktiske erfaringer	85
10.3. Konklusjon	86
Sammendrag på engelsk	87
Appendiks. Grunnbegreper og definisjoner	107
A.1. Notasjon	107
A.2. Målenivå og variabeltyper	108
A.3. Elementer av lineær algebra og matematisk statistikk	109
A.3.1. Definitte matriser	109
A.3.2. Kovariansmatriser	110
A.3.3. Den multinormale fordeling	110
A.3.4. Konvergens i sannsynlighet	111
A.3.5. Konvergens i fordeling	111
A.3.6. Informasjonsmatrisen	114
A.3.7. Cramér-Rao's ulikhet	115
A.3.8. Om asymptotisk optimalitet	116
A.3.9. Sannsynlighetsmaksimeringsprinsippet (ML)	117
Litteratur	118
Utkommet i serien Rapporter	121

Standardtegn

- . Tall kan ikke forekomme
- .. Oppgave mangler
- Null
- 0 Mindre enn 0,5 av den brukte enhet

1. Introduction	13
2. Background and purpose	14
3. Models for qualitative response variables	16
4. The linear model	20
4.1. Binary response	20
4.2. Multinary response	21
4.3. Estimation by two-stage generalized least squares (2SGLS)	25
4.4. The link between individual and grouped data	28
4.5. Estimation by ordinary least squares (OLS)	29
4.6. Estimation by maximum likelihood (ML)	31
4.7. Estimation by minimum Neyman chi-square ($\min x_N^2$)	32
4.8. Comparing OLS and 2SGLS	35
5. The log-linear model	35
5.1. Binary response	35
5.2. Multinary response	38
5.3. Estimation by two-stage generalized least squares (2SGLS)	39
5.4. Estimation by maximum likelihood (ML)	42
5.4.1. Multinomial logit model	42
5.4.2. Symmetric log-linear formulation	44
5.5. Hierarchical hypotheses for the log-linear model	44
5.6. Estimation by minimum logit chi-square ($\min x_L^2$)	47
5.7. Comparing 2SGLS and ML	48
6. The probit model	50
7. Goodness-of-fit measures for the linear and log-linear models	53
7.1. The coefficient of determination (R^2) in the linear model	53
7.1.1. On OLS estimation in general	53
7.1.2. On binary response regression	54
7.1.2.1. Theoretical discussion	54
7.1.2.2. Numerical examples	61
7.2. Chi-square measures	65
7.2.1. Pearson chi-square (x_p^2)	65
7.2.2. Neyman chi-square (x_N^2)	66
7.2.3. The likelihood ratio chi-square (G^2)	66
7.3. The index of misclassification (C)	66
7.4. Goodness-of-fit measures: a recapitulation	67
8. Empirical approach	67
8.1. Scope	67
8.2. The data set	67
8.3. The problem of empty cells	70
8.4. Software	71

CONTENTS (cont.)

	Page
9. Empirical results	71
9.1. Comparing goodness-of-fit	71
9.2. Variance estimates for the linear model. Comparing 2SGLS and OLS	74
9.3. Efficiency gain with respect to the saturated model	76
9.4. Hypothesis testing	78
9.4.1. Log-linear model	78
9.4.2. Linear model	80
9.4.3. Comparing the linear and the log-linear models	81
10. Attempt at a synthesis	81
10.1. A priori arguments	81
10.2. Practical experiences	85
10.3. Conclusion	86
Summary in English	87
Appendix. Basic concepts and definitions	107
A.1. Notation	107
A.2. Levels of measurement and variable typology	108
A.3. Elements of linear algebra and mathematical statistics	109
A.3.1. Definite matrices	109
A.3.2. Covariance matrices	110
A.3.3. The multivariate normal distribution	110
A.3.4. Convergence in probability	111
A.3.5. Convergence in law	111
A.3.6. The information matrix	114
A.3.7. The Cramér-Rao inequality	115
A.3.8. On asymptotic optimality	116
A.3.9. The maximum likelihood principle (ML)	117
Literature	118
Publications issued in the series Rapporteur	121

Explanation of Symbols

- . Category not applicable
- .. Data not available
- Nil
- 0 Less than 0.5 of
unit employed

FIGURREGISTER

	Side
4.1. Den lineære sannsynlighetsfunksjon	20
5.1. Logit-funksjonen	36
6.1. Probit-funksjonen	50
6.2. Sammenhengen mellom p_g og ϕ_g	52
7.1. Sannsynlighetene i eks. 1-6 opptegnet for $s = 24$	62
8.1. Histogrammer som viser hvordan de deltids- og heltidsarbeidende personene i utvalget for- deler seg etter den gjennomsnittlige deltids-, henh. heltidsandel blant personer i samme alders-, utdannings- og sivilstandsgruppe	69
9.1. Relativ effisiens av 2SGLS i forhold til separate cellegjennomsnitt, som funksjon av ut- valgsstørrelsen n_g . Linear modell uten samspillsledd, $\delta = 0,5$	77
10.1. Forløpet av den lineære, logit og probit sannsynlighetsfunksjonene i intervallet fra $\frac{1}{2}$ og oppover	82
10.2. Oversikt over noen mulige estimeringsmetoder for ulike sannsynlighetsmodeller med kvali- tativ avhengig variabel, med henvisning til avsnitt i notatet	84

TABELLREGISTER

	Side
2.1. Estimerte yrkesfrekvenser for personer ikke under utdanning, i grupper for kjønn, alder, ekteskapelig status og utdanningsnivå. Årsgjennomsnitt 1977. Prosent	14
2.2. Tallet på personer ikke under utdanning i AKU-utvalget for 1977, etter kjønn, alder, ekteskapelig status og utdanningsnivå. Totalt for alle 4 kvartal	15
3.1. Eksempel på binær-variable i en modell med 11 parametre	18
3.2. Oversikt over fler-faktor-effekter i en mettet modell med dikotom avhengig variabel	19
7.1. Talleksempler for sannsynlighetene p_{g1} ($g=1,2,\dots,s$) og formler for utregning av ω , σ^2 og plim R^{*2}	61
7.2. Eksempel 1. Verdien av τ_1/τ_2 og av plim R^{*2} , for utvalgte verdier av s og n	63
7.3. Eksempel 2. Verdien av τ_1/τ_2 og av plim R^{*2} , for utvalgte verdier av s og n	64
7.4. Eksempel 3 og 4. Verdien av τ_1/τ_2 og av plim R^{*2} , for utvalgte verdier av s og n	64
7.5. Eksempel 5. Verdien av τ_1/τ_2 og av plim R^{*2} , for utvalgte verdier av s og n	64
7.6. Eksempel 6. Verdien av τ_1/τ_2 og av plim R^{*2} , for utvalgte verdier av s og n	65
8.1. Datamaterialet. Menn 16-74 år, etter sivilstand, utdanningsnivå, alder og arbeidstid. Absolutte tall (n_{gj}) og prosent ($100 \cdot y_{gj}$)	68
9.1. Føyningsmål utregnet for $\delta = 0,01$	72
9.2. Føyningsmål utregnet for $\delta = 0,5$	72
9.3. Varians- og effisiensestimater for estimatorene b_{jk} (2SGLS) og \hat{b}_{jk} (OLS). Lineær modell uten samspillsledd, $\delta = 0,5$	74
9.4. Varians- og effisiensestimater for estimatorene b_{jk} (2SGLS) og \hat{b}_{jk} (OLS). Lineær modell med alle 1.-ordens samspillsledd, $\delta = 0,5$	74
9.5. Varians- og effisiensestimater for estimatorene $b_{ji} - b_{jk}$ (2SGLS) og $\hat{b}_{ji} - \hat{b}_{jk}$ (OLS), for utvalgte i, k . Lineær modell uten samspillsledd, $\delta = 0,5$	75
9.6. Testresultater for den log-lineære modell. Kji-kvadratobservatoren (G^2), antall frihetsgrader (ν_f) og signifikanssannsynlighet ($p(f)$), ved betinget og ubetinget testing. $\delta = 0$.	78
9.7. Testresultater for den log-lineære modell. Kji-kvadratobservatoren (G^2), antall frihetsgrader (ν_f) og signifikanssannsynlighet ($p(f)$), ved betinget og ubetinget testing. $\delta = 0,01$	79
9.8. Testresultater for den log-lineære modell. Kji-kvadratobservatoren (G^2), antall frihetsgrader (ν_f) og signifikanssannsynlighet ($p(f)$), ved betinget og ubetinget testing. $\delta = 0,5$	79
9.9. Testresultater for den lineære modell. Kji-kvadratobservatoren (x_N^2), antall frihetsgrader (ν_f) og signifikanssannsynlighet ($p(f)$), ved betinget og ubetinget testing. $\delta = 0,5$	80
9.10. Sammenlikning av signifikanssannsynligheter for tilsvarende tester i den lineære og den log-lineære modell	81

INDEX OF FIGURES

	Page
4.1. The linear probability function	20
5.1. The logit function	36
6.1. The probit function	50
6.2. The relationship between p_g and ϕ_g	52
7.1. The probabilities in ex. 1-6, illustrated for $s = 24$	62
8.1. Histograms showing the distribution of part-time and full-time workers in the sample, by the proportion of part-time and full-time workers, respectively, among respondents with identical values on the independent variables	69
9.1. Relative efficiency of 2SGLS as compared with crude cell means, as a function of the subsample size n_g . Linear model without interaction terms, $\delta = 0.5$	77
10.1. The linear, logit, and probit quantal response curves, for $p = 0.5$ and above	82
10.2. Survey of feasible methods of estimation for various qualitative response models, with reference to the relevant sections of this paper	84

INDEX OF TABLES

	Page
2.1. Estimated labour force participation rates for persons not under education, in groups for sex, age, marital status and level of education. Annual average 1977. Per cent	14
2.2. Persons not under education present in the labour force survey sample for 1977, by sex, age, marital status and level of education. Total for all 4 quarters	15
3.1. Example of dummy variables in a probability model with 11 parameters	18
3.2. Multi-factor effects in a saturated model with one dichotomous dependent variable	19
7.1. Numerical examples for the true probabilities p_{g1} ($g=1,2,\dots,s$) and formulae for calculating ω , σ^2 , and $\text{plim } R^{*2}$	61
7.2. Example 1. Value of τ_1/τ_2 and of $\text{plim } R^{*2}$, for selected values of s and n	63
7.3. Example 2. Value of τ_1/τ_2 and of $\text{plim } R^{*2}$, for selected values of s and n	64
7.4. Examples 3 and 4. Value of τ_1/τ_2 and of $\text{plim } R^{*2}$, for selected values of s and n	64
7.5. Example 5. Value of τ_1/τ_2 and of $\text{plim } R^{*2}$, for selected values of s and n	64
7.6. Example 6. Value of τ_1/τ_2 and of $\text{plim } R^{*2}$, for selected values of s and n	65
8.1. The data set. Men 16-74 years old, by marital status, level of education, age, and weekly number of hours worked. Absolute numbers (n_{gj}) and per cent ($100 \cdot y_{gj}$)	68
9.1. Goodness-of-fit measures calculated for $\delta = 0.01$	72
9.2. Goodness-of-fit measures calculated for $\delta = 0.5$	72
9.3. Variance and efficiency estimates for the estimators b_{jk} (2SGLS) and $\hat{\beta}_{jk}$ (OLS). Linear model without interaction terms, $\delta = 0.5$	74
9.4. Variance and efficiency estimates for the estimators b_{jk} (2SGLS) and $\hat{\beta}_{jk}$ (OLS). Linear model with all 1st order interaction terms, $\delta = 0.5$	74
9.5. Variance and efficiency estimates for the estimators $b_{ji} - b_{jk}$ (2SGLS) and $\hat{\beta}_{ji} - \hat{\beta}_{jk}$ (OLS), for selected i, k . Linear model without interaction terms, $\delta = 0.5$	75
9.6. Testing results for the log-linear model. The likelihood ratio chi-square statistic (G^2), degrees of freedom (ν_f), and p-value ($p(f)$), for conditional and unconditional testing. $\delta = 0$	78
9.7. Testing results for the log-linear model. The likelihood ratio chi-square statistic (G^2), degrees of freedom (ν_f), and p-value ($p(f)$), for conditional and unconditional testing. $\delta = 0.01$	79
9.8. Testing results for the log-linear model. The likelihood ratio chi-square statistics (G^2), degrees of freedom (ν_f), and p-value ($p(f)$), for conditional and unconditional testing. $\delta = 0.5$	79
9.9. Testing results for the linear model. The Neyman chi-square statistic (χ_N^2), degrees of freedom (ν_f), and p-value ($p(f)$), for conditional and unconditional testing. $\delta = 0.5$	80
9.10. Comparison of p-values for analogous tests in the linear and log-linear models	81

1. INNLEDNING OG DISPOSISJON

Formålet med denne rapporten er å studere enkelte matematisk-statistiske metoder. Vi sammenlikner ulike statistiske modeller for situasjoner der sannsynlighetsfordelingen for en kvalitativ avhengig variabel er bestemt av en rekke observerbare uavhengige variable. Slike modeller har vært gjenstand for stadig sterkere interesse i de senere år. Et område der slike modeller har anvendelse, er studier av individers yrkesdeltaking. Dette er også siktepunktet i vårt tilfelle, og vi har derfor benyttet data fra Statistisk Sentralbyrås Arbeidskraftundersøkelser (AKU) til den empiriske utprøving av metodene. Vekten i denne rapporten ligger imidlertid helt og holdent på metodesiden. Datamaterialet er ikke lagt til rette med sikte på å si noe om sammenhengene på arbeidsmarkedet og blir heller ikke brukt til dette, men kun til å illustrere beregningsmetodene.

Siktemålet med dette arbeidet er praktisk-empirisk snarere enn ren-teoretisk. Modeller og estimeringsteknikker som er vanskelige å gjennomføre i praksis, får liten plass. Til gjengjeld går vi nokså langt i retning av å spesifisere hvordan de enkelte modeller må formuleres for at parametrene skal kunne estimeres ved hjelp av de data og den EDB-utrustning som er tilgjengelig i Statistisk Sentralbyrå. Det er særlig tilgangen på standardprogrammer som her setter grenser.

Rapporten er redigert som følger:

I kapittel 2 gjør vi rede for ett motiv for å anvende sannsynlighetsmodeller for kvalitative avhengige variable. Kapittel 3 er en elementær gjennomgang av hovedprinsippene for konstruksjon av sannsynlighetsmodeller for kvalitative avhengige variable, slik vi anvender dem i denne rapporten. I kapitlene 4, 5 og 6 gjennomgår vi mer i detalj tre hovedtyper av modeller, viser hvordan parametrene kan estimeres og testes og vurderer de ulike metoder mot hverandre. Kapittel 7 behandler en del føyningsmål for sammenlikning av modeller og estimeringsmetoder. Kapitlene 8 og 9 inneholder en beskrivelse av den empiriske del av analysearbeidet og av de resultater vi finner. Oppsummering og konklusjon kommer i avsnitt 10.

I et appendiks har vi dessuten framstilt en del grunnbegreper og definisjoner som drøftingen i denne rapporten bygger på. Notasjonen i rapporten er således forklart i avsnitt A.1. Avsnitt A.2 redegjør for den terminologien som anvendes for å skille mellom ulike variabeltyper og målenivå. Avsnitt A.3 er en oversikt over det matematisk-statistiske verktøy som blir brukt i rapporten. Avsnittet kan være nyttig for lesere uten særlig grunnlag i matematisk statistikk og som likevel ønsker å følge hele utledningen.

Lesere som ønsker en mer kortfattet framstilling, kan nøye seg med det engelske sammendraget til slutt.

Rapporten er i første rekke et forsøk på en noenlunde systematisk sammenstilling av nokså velkjente resultater innenfor feltet. Spesielt er det lagt vekt på å beskrive den lineære og den log-lineære modellen langs nokså parallelle linjer. På enkelte punkter kan rapporten trolig likevel gi en viss ny innsikt. Dette gjelder særlig den empiriske sammenlikning av de ulike modelltyper og inferensmetoder.

Vanskelighetsgraden varierer. De fleste avsnittene forutsetter kjennskap til elementær matrisialgebra. Alle som har såpass bagasje skulle imidlertid, ved først å lese appendikset, kunne tråkle seg helt igjennom

Forfatteren skylder takk til en rekke kolleger for hjelp og veiledning i arbeidet med rapporten. Ivar Sand har gjort programmeringsarbeidet. Liv Hansen har tegnet figurene og gjort et stort manuelt beregningsarbeid. Per Sevaldson, Tor Haldorsen og Herdis Thorén Amundsen har, uten ethvert ansvar for mulige feil, lest manuskriptet og gitt særs verdifulle kommentarer. Tor Haldorsen og Erik Hernæs har dessuten stadig ytt råd og dåd underveis. Eva Sundby har maskinkrevet rapporten med største teft og tålmodighet.

2. BAKGRUNN OG FORMÅL

Den viktigste kilden til data om arbeidsmarkedet i Norge er Statistisk Sentralbyrås Arbeidskraftundersøkelser (AKU). Hvert kvartal intervjues ca. 10 000 tilfeldig utvalgte personer mellom 16 og 74 år om sitt forhold til arbeidsmarkedet. AKU er den eneste kilde til statistikk om arbeidsmarkedet som omfatter personer med alle typer arbeidsmarkedstilknytning: lønnstakere, selvstendige yrkesutøvere, familiearbeidskraft, midlertidig fraværende, arbeidsledige og personer utenfor arbeidsstyrken (husmødre, vernepliktige, pensjonister o.a.). Dette er av vesentlig betydning dersom formålet er å klarlegge hvilke faktorer som påvirker personenes arbeidsmarkedstilknytning.

Svakhetene ved AKU henger i første rekke sammen med at det er en utvalgsundersøkelse med et begrenset utvalg. Alle tall er beheftet med utvalgsfeil. Denne er relativt større jo mindre den persongruppen er som tallet gjelder. Ifølge AKU var det f.eks. i 1977 om lag 6 000 gifte kvinner som var arbeidssøkere uten arbeidsinntekt, regnet i gjennomsnitt over året. Dette tallet er imidlertid estimert med et standardavvik på anslagsvis hele 1 000 personer (se Statistisk Sentralbyrå (1978)).

Dette betyr at dersom vi kryssfordeler befolkningen etter en rekke kjennetegn, f.eks. kjønn, alder, ekteskkelig status, utdanning og region, og ønsker å studere arbeidsmarkedsatferden til alle disse gruppene hver for seg, så blir AKU-materialet raskt for spinkelt. Mange av gruppene blir for små til at AKU kan gi pålitelige estimater.

Statistisk Sentralbyrå har ved flere anledninger laget framskrivinger av arbeidsstyrken basert nettopp på en slik oppsplitting av befolkningen, se Hernæs et al. (1977) og Fridstrøm (1978). Ved den siste framskrivingen i 1978 tok en f.eks. utgangspunkt i følgende tabell, estimert på grunnlag av AKU 1977:

Tabell 2.1. Estimerte yrkesfrekvenser for personer ikke under utdanning, i grupper for kjønn, alder, ekteskkelig status og utdanningsnivå. Årsgjennomsnitt 1977. Prosent *Estimated labour force participation rates for persons not under education in groups for sex, age, marital status and level of education. Annual average 1977. Per cent*

Kjønn, ekteskkelig status og utdanningsnivå (I og II) <i>Sex, marital status and level of education (I or II)</i>	Alder. År <i>Age. Years</i>							
	16-19	20-24	25-29	30-39	40-49	50-59	60-66	67-74
Menn <i>Men</i>								
I	89	73	91	95	93	86	68	25
II	93	80	99	99	99	96	81	35
Ugifte kvinner <i>Unmarried women</i>								
I	75	74	47	64	51	51	44	13
II	85	91	86	92	92	88	66	22
Gifte og før gifte kvinner <i>Married and previously married women</i>								
I	61	34	37	56	65	50	34	8
II	54	56	61	64	69	65	44	12

Kilde: Arbeidskraftundersøkelsene (upublisert materiale i Byrået).

Source: The labour force sample surveys (unpublished data).

Tabell 2.1 viser yrkesprosentene for en del ulike befolkningsgrupper i 1977. Med yrkesprosent mener vi her prosentandelen av alle personer i hver gruppe som er i arbeidsstyrken, slik dette begrepet er definert i AKU. Med "utdanningsnivå I" mener vi bare grunnutdanning, og med "utdanningsnivå II" mener vi all utdanning utover grunnutdanning.

Estimatene i tabell 2.1 er utregnet ved å dele opp befolkningen ikke under utdanning i 48 grupper eller "celler", nemlig de som framkommer ved kryssgruppering av kjennetegnene i forspalte og tabellhode. For hver celle har en deretter regnet ut prosentandelen i utvalget som er i arbeidsstyrken - en finner gjennomsnittlig antall yrkesaktive pr. 100 personer i cellen.

Grunnlaget for utregning av gjennomsnitt varierer imidlertid fra celle til celle. I noen grupper er utvalget svært spinkelt - dette framgår klart av tabell 2.2. Flere av yrkesprosentene for ugifte kvinner bygger f.eks. på utvalg av færre enn 100 personer. Det sier seg selv at disse estimatene er svært usikre.

Tabell 2.2. Tallet på personer ikke under utdanning i AKU-utvalget for 1977, etter kjønn, alder, ekteskapelig status og utdanningsnivå. Totalt for alle 4 kvartal *Persons not under education present in the labour force survey sample for 1977, by sex, age, marital status and level of education. Total for all 4 quarters*

Kjønn, ekteskapelig status og utdanningsnivå (I og II) <i>Sex, marital status and level of education (I or II)</i>	Alder. År <i>Age. Years</i>							
	16-19	20-24	25-29	30-39	40-49	50-59	60-66	67-74
Menn <i>Men</i>								
I	326	386	432	930	1 039	1 488	1 133	1 126
II	375	1 303	1 294	2 605	1 860	2 030	1 121	759
Ugifte kvinner <i>Unmarried women</i>								
I	249	145	57	85	45	131	138	223
II	389	690	208	166	77	97	112	121
Gifte og før gifte kvinner <i>Married and previously married women</i>								
I	18	201	355	1 047	1 267	1 882	1 346	1 108
II	21	491	1 076	2 374	1 590	1 507	810	516

K i l d e: Arbeidskraftundersøkelsene (upubliseret materiale i Byrået).

Source: *The labour force sample surveys (unpublished data).*

Grupperingen i tabell 2.1 og 2.2 er ikke spesielt fin. Når formålet er å lage framskrivninger av arbeidskrafttilbudet, kan det f.eks. være ønskelig å benytte en langt finere gruppering. Først og fremst er det i arbeidsmarkedssammenheng behov for å kunne dele inn befolkningen i langt flere enn to utdanningskategorier. Det er dessuten en rekke andre variable som har betydning for arbeidstilbudet, enn de som inngår i tabell 2.1. Ett eksempel på dette er tallet på barn og alder på barna i husholdningen. Særlig for gifte kvinner har dette vist seg å ha stor betydning for yrkesdeltakingen (Se Ljones (1979)).

Dersom vi skal kunne estimere yrkesprosenten i en enda finere oppdelt bestand enn tabell 2.1, må vi benytte andre metoder enn å beregne enkle cellegjennomsnitt i tabellen. Et blick på tabell 2.1 viser at det er sammenheng - "struktur" - mellom de ulike cellene i tabellen, slik at en god del informasjon går tapt når vi estimerer hver celle for seg. Dette poenget er blitt formulert særlig klart av A. Amundsen (1968), s. 11-12:

Metoden [enkle cellegjennomsnitt] innebærer, kan en si, at materialet splittes opp i et større antall del-materialer, som så behandles som atskilte undersøkelser. Tallet på del-undersøkelser er lik tallet på ruter i tabellen hvor det forekommer tall..... For hver del-undersøkelse regnes ut et resultat..... Dette resultat regnes ut uten hensyn til hvilke resultater som kommer fram i tabellens øvrige ruter. Ved å gå fram på denne måten unnlater vi å ta hensyn til en særdeles viktig del av de opplysninger materialet inneholder, nemlig at variasjonen fra rute til rute innenfor en linje og innenfor en kolonne i tabellen har en bestemt tendens. (Amundsens utvevelser.)

I vårt materiale kan vi iaktta bl.a. følgende "struktur" eller "tendenser": Yrkesprosentene stiger nokså entydig med alderen, inntil en viss alder, da de begynner å synke. For menn er det unntak fra dette mønsteret i aldersgruppen med mange vernepliktige, for kvinner i aldersgruppene med høy fruktbarhet eller stor barneomsorgsbyrde. Yrkesdeltakingen stiger med utdanningsnivået, og ugifte kvinner har høyere yrkesdeltaking enn gifte.

Ved å bruke en estimeringsteknikk som utnytter denne informasjonen, må det være mulig å estimere den samme tabellen med større grad av presisjon enn før. Alternativt kan en estimere en mye finere oppdelt tabell uten å øke usikkerheten i særlig grad.

3. MODELLER FOR KVALITATIVE AVHENGIGE VARIABLE

Tabell 2.1 er dannet ved separat utregning av 48 cellegjennomsnitt. Sagt på en annen måte, yrkesprosenttabellen er estimert som funksjon av 48 parametre - én for hver rute i tabellen. Det å utnytte strukturen i materialet innebærer at vi vil estimere de samme 48 tallene ved hjelp av langt færre (m) parametre. Hver yrkesprosent y_g i tabellen anses da å være en gitt funksjon h_g av disse parametrene:

$$(3.1) \quad y_g = h_g(\beta_1, \beta_2, \dots, \beta_m) + u_g \quad (g=1, \dots, 48),$$

idet vi kaller parametrene $\beta_1, \beta_2, \dots, \beta_m$. u_g er et stokastisk feilledd med forventning null. Når funksjonene h_g er gitt, består oppgaven i å estimere $\beta_1, \beta_2, \dots, \beta_m$. Alle parametre blir estimert simultant, det vil si at hver celle i tabellen i prinsippet bidrar informasjon til estimeringen av hver parameter. La oss betegne estimatene $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_m$. Etter at vi har funnet disse, kan vi regne oss tilbake til estimater \hat{y}_g av forventningen til y_g - $E(y_g)$ - ved hjelp av relasjonen

$$(3.2) \quad \hat{y}_g = h_g(\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_m).$$

"Forventningen til y_g " er den verdi vi ville finne hvis vi talte opp alle medlemmer av gruppe g i populasjonen. Det er denne størrelsen vi egentlig er interessert i å fastslå. I et tilfeldig utvalg fra persongruppe g vil y_g være et forventningsrett estimat, beheftet med utvalgsfeil, av populasjonsstørrelsen $E(y_g)$. Et annet estimat av $E(y_g)$ er altså \hat{y}_g , hvis fordelingsegenskaper avhenger av funksjonsformen h_g , fordelingen til restleddene u_g og estimeringsmetoden for $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_m$.

I denne rapporten vil vi innskrenke oss til å se på en noe mindre generell formulering enn (3.1), nemlig

$$(3.3) \quad y_g = h(x_g' \beta) + u_g$$

der $\beta = (\beta_1, \beta_2, \dots, \beta_m)'$ er vektoren av parametre og $x_g = (x_{g1}, x_{g2}, \dots, x_{gm})'$ er en $(m \times 1)$ vektor av kjennetegn for individene i gruppe g . Funksjonen h er her altså den samme for alle g , og argumentet er en lineærkombinasjon $x_g' \beta = \sum_{j=1}^m x_{gj} \beta_j$ av parametrene, der vektene x_{gj} varierer med g . Kjennetegnvektoren x_g kan i prinsippet bestå av alle typer variable, f.eks. også konstanter eller produktledd.

Modellen (3.3) representerer en bedre estimeringsmetode enn enkle cellegjennomsnitt bare dersom funksjonsformen h og estimeringsmetoden for β er fornuftig valgt. Vi skal i dette notatet særlig ta for oss to vanlige spesifikasjoner av h , nemlig den lineære og den log-lineære funksjon. For hver av modellene vil vi betrakte noen mulige estimeringsmetoder. Framstillingen i dette avsnittet vil imidlertid gjelde for generell h .

Dersom en bruker modellen (3.3), er det naturlig å se på hvert individ i i utvalget som realisasjonen av en binær (Bernoulli) stokastisk variabel y_i som er 1 dersom individ i er i arbeidsstyrken, og 0 ellers:

$$(3.4) \quad \begin{cases} P(y_i=1) = h(x_i' \beta) \\ P(y_i=0) = 1 - h(x_i' \beta). \end{cases}$$

Individ i er i arbeidsstyrken med sannsynlighet $h(x_i'\beta)$ og utenfor arbeidsstyrken med sannsynlighet $1 - h(x_i'\beta)$. x_i er kjennetegnvektoren til individ i . Alle individer med samme verdi på vektoren x_i forutsettes altså å ha samme sannsynlighet for å være i arbeidsstyrken.

Ekvivalent med (3.4) kan vi skrive

$$(3.5) \quad y_i = h(x_i'\beta) + u_i,$$

der u_i representerer avviket mellom realisert og forventet verdi på y_i . u_i er en stokastisk variabel med fordeling gitt ved

$$(3.6) \quad \begin{cases} P(u_i = 1 - h(x_i'\beta)) = h(x_i'\beta) \\ P(u_i = -h(x_i'\beta)) = 1 - h(x_i'\beta). \end{cases}$$

Vi har altså

$$(3.7) \quad E(u_i) = 0, \quad \text{var}(u_i) = h(x_i'\beta)(1 - h(x_i'\beta)).$$

En kan velge om en her vil se på vektoren x_i som stokastisk eller ikke-stokastisk. I det første tilfellet må vi betrakte fordelingen (3.6) som betinget med hensyn på x_i . Dette vil være underforstått i det følgende.

Hva er forbindelsen mellom modell (3.5) og modell (3.3)? La oss tenke oss at vi i vårt utvalg har et antall n_g individer med samme kjennetegnvektor x_i , altså

$$(3.8) \quad x_i = x_g \quad \text{for } n_g \text{ individer } i.$$

La oss videre forutsette at de enkelte individenes valg av arbeidsstyrkestatus y_i er ukorrelert med hverandre, dvs. $E(u_i u_j) = 0$ for $i \neq j$. Variabelen

$$(3.9) \quad n_g y_g = \sum_{i=1}^{n_g} y_i$$

er da binomisk fordelt med parametre n_g og $h(x_g'\beta)$. Vi har dermed

$$(3.10) \quad E(y_g) = h(x_g'\beta), \quad E(u_g) = 0$$

$$(3.11) \quad \text{var}(y_g) = \text{var}(u_g) = \frac{1}{n_g} h(x_g'\beta)(1 - h(x_g'\beta))$$

Relasjonen (3.3), som beskriver fordelingen til en gruppe av individer, framkommer altså ved at vi regner gjennomsnittet

$$(3.12) \quad y_g = \frac{1}{n_g} \sum_{i=1}^{n_g} y_i$$

for alle de individer hvis fordeling er gitt av relasjon (3.5).

La oss her forsøke å trekke trådene tilbake til det eksemplet som dannet utgangspunkt for diskusjonen. y_g ($g=1, \dots, 48$) svarer til yrkesprosentene i tabell 2.1. n_g ($g=1, \dots, 48$) svarer på samme måte til tallene i tabell 2.2. x_g kan i dette tilfelle tenkes å bestå av 4 sett med binær-variable ("dummy"-variable) x_{gj} , ett sett for hver variabel i tabellhode eller forspalte, pluss et konstantledd, f.eks. slik som i tabell 3.1.

Tabell 3.1. Eksempel på binær-variable i en modell med 11 parametre *Example of dummy variables in a probability model with 11 parameters*

Variabel <i>Variable</i>	Verdi <i>Value</i>	Binærvariable <i>Dummy variables</i>
Konstant <i>Constant</i>		$x_{g1} = 1$
Kjønn <i>Sex</i>	Mann <i>Male</i>	$x_{g2} = 0$
	Kvinne <i>Female</i>	" = 1
Alder <i>Age</i>	16-19 år/years	$x_{g3}, \dots, x_{g9} = 0,0,0,0,0,0,0$
	20-24 " "	" = 1,0,0,0,0,0,0
	25-29 " "	" = 0,1,0,0,0,0,0
	30-39 " "	" = 0,0,1,0,0,0,0
	40-49 " "	" = 0,0,0,1,0,0,0
	50-59 " "	" = 0,0,0,0,1,0,0
	60-66 " "	" = 0,0,0,0,0,1,0
Ekteskadelig status <i>Marital status</i>	Menn, ugifte kvinner <i>Men, unmarried women</i>	$x_{g10} = 0$
	Gifte og før gifte kvinner <i>Married and previously married women</i>	" = 1
Utdanningsnivå <i>Level of education</i>	I	$x_{g11} = 0$
	II	" = 1

Når vi beskriver en kvalitativ (gruppeinndelt) variabel ved et sett av binærvariable, får vi en binærvariable for hver verdi den opprinnelige variable kan anta, minus én. Den siste kalles "basisverdien" eller "basisgruppen". Hvis den opprinnelige variable antar basisverdien, får alle binærvariable verdi 0. Hvis den opprinnelige variable har en annen verdi, får den tilsvarende binærvariable verdi 1 og de øvrige får verdi 0.

Betrakt tabell 3.1. Variabelen kjønn har to mulige verdier og beskrives entydig av én eneste binærvariable (x_{g2}), som antar verdien 1 hvis utvalgspersonen er kvinne, 0 ellers. Variabelen alder har 8 mulige verdier og beskrives av et sett av 7 binærvariable. Basisgruppen er 16-19 år. Variabelen x_{g3} er 1 hvis og bare hvis utvalgspersonen er 20-24 år, osv.

Hele vektoren x_g består av alle 11 binærvariable x_{g1}, \dots, x_{g11} . For gruppen av ugifte kvinner i alder 30-39 år og med utdanningsnivå II ser f.eks. x_g slik ut:

$$(3.13) \quad x_g = (1 \ 1 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1)'$$

Tilsvarende kjennetegnvektoren x_g har vi en parametervektor $\beta = (\beta_1, \dots, \beta_{11})'$. Hver parameter β_j (unntatt konstantleddet β_1) er et uttrykk for avvik i forhold til basisgruppen.

Den presise tolkning av parametrene vil imidlertid avhenge av funksjonsformen h . Tallet $h(\beta_1)$ vil angi yrkesdeltakingssannsynligheten for en person med basisverdi på alle de 4 variable, dvs. menn i alder 16-19 år og med utdanningsnivå I; for disse er

$$(3.14) \quad x_g = (1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0)'$$

$$\text{og således } \sum_{j=1}^{11} x_{gj} \beta_j = \beta_1.$$

Denne formuleringen innebærer en antakelse om at vi kan uttrykke de 48 sannsynlighetene tilsvarende tabell 2.1 som er funksjon av i alt bare 11 parametre. De 11 parametrene som er tatt med her, kan vi kalle de direkte virkningene mellom hver av de uavhengige variable og den avhengige variable. Alternativt kan vi snakke om to-faktor-effekter. Hver parameter uttrykker nemlig sammenhengen mellom bare to faktorer: den avhengige variable (yrkesdeltaking) og én av de uavhengige variable (henholdsvis kjønn, alder, sivilstand eller utdanning).

Estimering i denne modellen ville gi helt andre estimater enn de som er gjengitt i tabell 2.1. Dersom vi tror at en modell med bare to-faktor-effekter gir en altfor grov beskrivelse av den "sanne" struktur, kan vi innføre såkalte samspillsledd, eller tre-faktor-effekter. Hvert samspillsledd vil her

være definert som produktet av to binærvariable fra ulike sett. Samspillsleddet uttrykker sammenhengen mellom tre variable: den avhengige variable og to uavhengige variable. Tankegangen bak et slikt samspill er at effekten av én uavhengig variabel varierer med nivået på en annen uavhengig variabel, f.eks. at utdanningsnivået ikke har like stor betydning for yrkesdeltakingen for menn som for kvinner.

Kan hende er vi av den oppfatning at også tre-faktormodellen er for grov. I såfall kan vi gå enda et skritt videre og operere med 2.-ordens samspill, eller fire-faktor-effekter, osv. I en situasjon med t variable kan vi i prinsippet definere helt opp til t-faktor-effekter. t-variabelmodellen med alle t-faktor- og lavere grads effekter kaller vi den mettede modell. En slik modell vil ha like mange parametre som det er sannsynligheter å estimere. Den mettede modellen har imidlertid i praksis lite for seg. Den vil gi de samme estimater som om vi estimerer hver celle for seg og innebærer således at vi ikke har lagt noen struktur på datamaterialet i det hele tatt.

Tabell 3.2. Oversikt over fler-faktor-effekter i en mettet modell med dikotom avhengig variabel
Multi-factor effects in a saturated model with one dichotomous dependent variable

Uavhengige variable <i>Independent variables</i>	Antall grupper <i>No. of groups</i>	Antall parametre <i>No. of parameters</i>	Sum antall parametre <i>Total no. of parameters</i>
Konstant <i>Constant</i>	1	<u>1</u>	1 (1)
2-faktor-effekter <i>2-factor effects</i>			
K	2	1	
A	8	7	
E	2	1	
U	2	<u>1</u>	10 (10)
3-faktor-effekter <i>3-factor effects</i>			
KxA	2x8	7	
KxE	2x2	1 (0)	
KxU	2x2	1	
AxE	8x2	7	
AxU	8x2	7	
ExU	2x2	<u>1</u>	24 (23)
4-faktor-effekter <i>4-factor effects</i>			
KxAxE	2x8x2	7 (0)	
KxAxU	2x8x2	7	
KxExU	2x2x2	1 (0)	
AxExU	8x2x2	<u>7</u>	22 (14)
5-faktor-effekter <i>5-factor effects</i>			
KxAxExU	2x8x2x2	7 (0)	<u>7 (0)</u> 64 (48)

Forklaring: K = kjønn, A = alder, E = ekteskapelig status, U = utdanningsnivå.

Legend: K = sex, A = age, E = marital status, U = level of education.

I tabell 3.2 har vi satt opp en oversikt over de fler-faktor-effekter som vil være aktuelle i vår modell. Tallene utenfor parentes gjelder egentlig for et materiale der også mennene tenkes fordelt etter sivilstand, dette gir i alt 64 grupper og følgelig 64 parametre i den mettede modell. I vårt materiale gjelder imidlertid sivilstandsgrupperingen bare for kvinner. Dette betyr at alle 1. og høyere ordens samspillsledd mellom kjønn og sivilstand er overflødige og ikke kan identifiseres. Dette bringer tallet på parametre i den mettede modell ned i 48, jfr. tallene i parentes. Estimering av denne modellen vil gi perfekt føyning til tabell 2.1, det vil si at innsetting av estimatene $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_{48}$ i relasjonen

$$(3.15) \quad \hat{y}_g = h(x_g' \hat{\beta}) \quad (g=1,2,\dots,48)$$

gir samme yrkesprosentestimer som gjengitt i tabellen. Dette vil gjelde uansett hvilken funksjonsform h som anvendes.

Hovedverket om statistisk analyse av kvalitative avhengige variable er Bishop et al. (1975). Cox (1970) er dessuten meget nyttig, men dekker ikke tilfellet med flere enn to mulige verdier på den avhengige variable. Mer kortfattede oversiktsartikler finner en hos Amemiya (1975) og McFadden (1976). McFadden (1974) og Domencich og McFadden (1975) gir bidrag til teorien og illustrerer samtidig modellenes anvendelighet på empiriske materialer.

4. DEN LINEÆRE SANNSYNLIGHETSMODELL

4.1 Binær avhengig variabel

Med utgangspunkt i modellen

$$(4.1) \quad y_g = h(x_g' \beta) + u_g$$

vil vi i dette avsnittet betrakte spesialtilfellet gitt ved funksjonen

$$(4.2) \quad h(x_g' \beta) = x_g' \beta.$$

Definer

$$(4.3) \quad p_g = E(y_g).$$

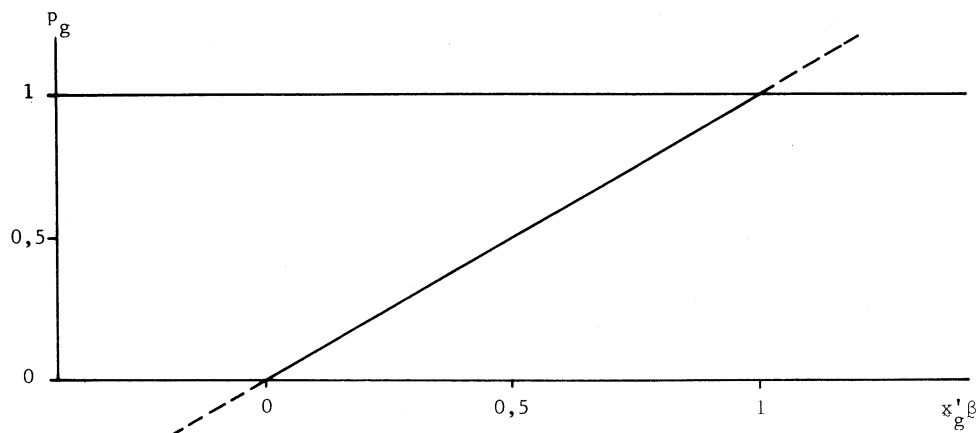
Da er altså

$$(4.4) \quad p_g = h(x_g' \beta) = x_g' \beta.$$

Sannsynligheten for at y_i skal anta verdien 1 er altså en lineær funksjon av $(\beta_1, \dots, \beta_m)$. Modellen (4.4) kaller vi derfor den lineære modell.

Funksjonen (4.4) er framstilt grafisk i figur 4.1.

Figur 4.1. Den lineære sannsynlighetsfunksjonen *The linear probability function*



Gitt at vi ønsker å tolke p_g som en sannsynlighet, byr spesifikasjonen (4.4) på visse prinsipielle problemer. Vi må da nemlig kreve at argumentet $x'_{g\beta}$ alltid ligger i intervallet $[0,1]$ - for andre verdier av $x'_{g\beta}$ er sannsynligheten p_g ikke definert. En alternativ modell-spesifikasjon til (4.4) kan være

$$(4.5) \quad p_g = h(x'_{g\beta}) = \begin{cases} 0 & \text{for } x'_{g\beta} < 0 \\ x'_{g\beta} & \text{for } 0 \leq x'_{g\beta} \leq 1 \\ 1 & \text{for } x'_{g\beta} > 1 \end{cases}$$

p_g blir da lineært stigende i intervallet $[0,1]$, og konstant lik 0 eller 1 på utsiden av intervallet.

En grei omtale av denne type modeller kan en finne hos Goldfeld og Quandt (1972), kap. 4.

Ljones (1979) har anvendt modellen i praksis til analyse av gifte kvinners yrkesdeltaking.

4.2 Multinær avhengig vektor

Til nå har vi betraktet situasjonen der den avhengige variable kan anta bare to verdier. Vi sier at variabelen er dikotom. Det har vært hensiktsmessig å "kalle" disse to verdiene 0 og 1. En slik variabel har vi kalt en binær eller dummy variabel. Ved å legge sammen flere uavhengige binærvariable med samme forventning får vi en binomisk variabel.

Det er lett å tenke seg tilfellet med to utfall generalisert til en variabel med r mulige, gjensidige utelukkende utfall, hvor r er et vilkårlig endelig positivt heltall. Vi kan tenke oss de mulige utfallene nummerert fra 1 til r , og kan da, for individ $i = 1, 2, \dots$, skrive

$$(4.6) \quad p_{ij} = P(v_i = j) \quad (j=1, 2, \dots, r)$$

Variabelen v_i , med sannsynlighetsfordeling gitt av (4.6), er en polytom variabel, som kan anta verdiene fra 1 til r .

For mange formål er det lettest å arbeide med en transformasjon av v_i , der vi gjør v_i om til en vektor av dummy-variable $y_{i.}^* = (y_{i1}^* \ y_{i2}^* \ \dots \ y_{ir}^*)'$. Hvert element i vektoren $y_{i.}^*$ er da gitt ved

$$(4.7) \quad y_{ij}^* = \begin{cases} 1 & \text{hvis } v_i = j \\ 0 & \text{ellers.} \end{cases} \quad (i=1, 2, \dots; j=1, 2, \dots, r)$$

Dersom v_i antar verdien j , er j -te element i $y_{i.}^*$ lik 1, de andre elementene er 0. En slik vektor kaller vi en multinær vektor.

Anta at det fins n_g individer i med samme sannsynlighetsfordeling, dvs. $p_{ij} = p_{gj}$ for alle j utfall og for alle n_g individer (i) som tilhører denne gruppen (g). Dersom individobservasjonene er stokastisk uavhengige, er summen av de n_g vektorene $y_{i.}^*$ tilordnet individene i gruppe g en multinomisk stokastisk vektor. Definer

$$(4.8) \quad y_{gj} = \frac{1}{n_g} \sum_i y_{ij}^*$$

der summasjonen går over alle individer i i gruppe g . Den multinomiske vektor er $(n_g y_{g1} \ n_g y_{g2} \ \dots \ n_g y_{gr})'$. Vi har her

$$(4.9) \quad p_{gj} = E(y_{gj}) = E(y_{ij}^*) = p_{ij}$$

for alle individer i i gruppe g .

For at (4.9) skal være en sannsynlighetsmodell, må vi ha

$$(4.10a) \quad \sum_{j=1}^r p_{gj} = 1 \quad (g=1,2,\dots,s)$$

$$(4.10b) \quad 0 \leq p_{gj} \leq 1 \quad (g=1,2,\dots,s; j=1,2,\dots,r).$$

Betingelsen (4.10a) innebærer at en polytom variabel med r mulige utfall er fullstendig beskrevet av $r-1$ binærvariable. Den r -te sannsynligheten får vi ved å trekke summen av de $r-1$ første fra 1. Dette prinsippet benyttet vi oss av allerede i kapittel 3, da vi brukte en multinær vektor bestående av 7 binær-variable til å fange opp den polytome variabelen "alder", med 8 mulige utfall.

Analogt med binærtifellet i avsnitt 4.1 kan vi nå postulere

$$(4.11) \quad p_{gj} = h(x_g' \beta_j) = x_g' \beta_j \quad (g=1,2,\dots,s; j=1,2,\dots,r).$$

Sannsynligheten for utfall j antas å være en lineærkombinasjon av en rekke parametre $\beta_j = (\beta_{j1} \beta_{j2} \dots \beta_{jm})'$.

Betingelsen (4.10b) volder atskillig større besvær i dette tilfellet enn når vi har bare to utfall å holde rede på. Det er vanskelig å tenke seg en enkel og fornuftig regel som sørger for at (4.10a) og (4.10b) er oppfylt samtidig, dersom $x_g' \beta_j$ faller utenfor intervallet $[0,1]$ for en eller flere g . Dette må anses å være en alvorlig innvending mot den lineære sannsynlighetsmodell for multinære vektorer.

I det generelle multinære tilfellet har y_{gj} -variablene følgende fordelingsegenskaper:

$$(4.12) \quad y_{gj} = x_g' \beta_j + u_{gj}$$

$$(4.13) \quad E(u_{gj}) = 0$$

$$(4.14) \quad \text{var}(u_{gj}) = \frac{1}{n_g} p_{gj}(1-p_{gj})$$

$$\left. \begin{array}{l} (g=1,2,\dots,s; \\ j=1,2,\dots,r) \end{array} \right\}$$

$$(4.15) \quad \text{cov}(u_{gj}, u_{hk}) = \begin{cases} 0 & \text{for } g \neq h \\ -\frac{1}{n_g} p_{gj} p_{gk} & \text{for } g = h, j \neq k \end{cases}$$

Restleddet u_{gj} er her det (stokastiske) avviket mellom y_{gj} og dens forventede verdi $p_{gj} = x_g' \beta_j$ (jfr.(3.5)). Restleddene for ulike grupper g og h forutsettes å være uavhengige, men restleddene tilhørende to ulike utfall j og k for samme gruppe er selvsagt korrelerte. n_g er antall individer i gruppe g . I spesialtilfellet $n_g = 1$ gjelder modellen som individmodell.

Definer matrisen

$$(4.16) \quad X = \begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_s' \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{s1} & x_{s2} & \dots & x_{sm} \end{bmatrix}$$

og vektorene

$$(4.17) \quad y_{\cdot j} = (y_{1j} \ y_{2j} \ \dots \ y_{sj})'$$

$$(4.18) \quad u_{\cdot j} = (u_{1j} \ u_{2j} \ \dots \ u_{sj})'$$

$$\left. \begin{array}{l} (j=1,2,\dots,r) \end{array} \right\}$$

Da kan vi skrive (4.12) i vektornotasjon som

$$(4.19) \quad y_{\cdot j} = X \beta_j + u_{\cdot j} \quad (j=1,2,\dots,r)$$

eller, enda mer kompakt,

$$(4.20) \quad \underline{y} = Z \underline{\beta} + \underline{u}$$

der

$$(4.21) \quad \underline{y} = (y'_{\cdot 1} \ y'_{\cdot 2} \ \dots \ y'_{\cdot r-1})'$$

$$(4.22) \quad Z = [I_{r-1} \otimes X] = \begin{bmatrix} X & \emptyset & \dots & \emptyset \\ \emptyset & X & \dots & \emptyset \\ \vdots & \vdots & & \vdots \\ \emptyset & \emptyset & \dots & X \end{bmatrix}$$

$$(4.23) \quad \underline{\beta} = (\beta'_1 \ \beta'_2 \ \dots \ \beta'_{r-1})'$$

$$(4.24) \quad \underline{u} = (u'_{\cdot 1} \ u'_{\cdot 2} \ \dots \ u'_{\cdot r-1})'$$

Restleddsvektoren \underline{u} i (4.20) har følgende fordelingsegenskaper:

$$(4.25) \quad E(\underline{u}) = 0$$

$$(4.26) \quad \Sigma = \text{Cov}(\underline{u}) = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \dots & \Sigma_{1,r-1} \\ \Sigma_{12} & \Sigma_{22} & \dots & \Sigma_{2,r-1} \\ \vdots & \vdots & & \vdots \\ \Sigma_{1,r-1} & \Sigma_{2,r-1} & \dots & \Sigma_{r-1,r-1} \end{bmatrix},$$

der

$$(4.27) \quad \Sigma_{jj} = \begin{bmatrix} p_{1j}(1-p_{1j})/n_1 & 0 & \dots & 0 \\ 0 & p_{2j}(1-p_{2j})/n_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & p_{sj}(1-p_{sj})/n_s \end{bmatrix} \quad (j=1,2,\dots,r-1)$$

og

$$(4.28) \quad \Sigma_{jk} = \begin{bmatrix} -p_{1j}p_{1k}/n_1 & 0 & \dots & 0 \\ 0 & -p_{2j}p_{2k}/n_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & -p_{sj}p_{sk}/n_s \end{bmatrix} \quad (j < k).$$

Kovariansmatrisen til \underline{u} , Σ , består altså av $(r-1)^2$ delmatriser eller blokker. Blokkene på diagonalen uttrykker variansene og blokkene utenfor diagonalen uttrykker kovariansene. Alle blokkene er dessuten selv diagonale matriser. Blokkene under hoveddiagonalen er derfor identiske med de tilsvarende blokker over hoveddiagonalen, og hele kovariansmatrisen Σ er selvsagt symmetrisk.

Selv om systemet (4.20) bare inneholder data og parametre for de $r-1$ første av de r mulige utfall, gir det en fullstendig beskrivelse av den simultane fordelingen til alle y_{gj} ($g=1,2,\dots,s$; $j=1,2,\dots,r$), siden, jfr. (4.10a),

$$(4.29) \quad \sum_{j=1}^r y_{gj} = 1 \quad (g=1,2,\dots,s).$$

La oss dessuten innføre følgende notasjon, som vil komme til nytte under utledningen i de påfølgende avsnitt.

Definer matrisene

$$(4.30) \quad N = \begin{bmatrix} n_1 & 0 & \dots & 0 \\ 0 & n_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & n_s \end{bmatrix}$$

$$(4.31) \quad M = I_{r-1} \otimes N$$

$$(4.32) \quad \Sigma_{ojk} = N \Sigma_{jk} \quad (j \leq k)$$

og

$$(4.33) \quad \Sigma_0 = M \Sigma.$$

Merk at matrisene Σ_{ojk} og Σ_0 er uavhengige av utvalgsstørrelsene n_1, \dots, n_s og består av variansene $p_{gj}(1-p_{gj})$ og kovariansene $-p_{gj}p_{gk}$ som gjelder når $n_g = 1$.

Definer dessuten de diagonale matrisene

$$(4.34) \quad P_n = \frac{1}{n} N$$

$$(4.35) \quad Q_n = I_{r-1} \otimes P_n = \frac{1}{n} M.$$

Diagonalelementene i P_n , n_g/n , angir da hvor stor andel gruppe g har av det totale utvalget. Vi vil anta at P_n og Q_n går mot grenseverdier gitt ved

$$(4.36) \quad \lim_{n \rightarrow \infty} P_n = P$$

$$(4.37) \quad \lim_{n \rightarrow \infty} Q_n = Q = I_{r-1} \otimes P$$

når utvalget vokser, altså at andelen n_g/n går mot en bestemt konstant.

Vi vil få bruk for relasjonene

$$(4.38) \quad \Sigma = (nQ_n)^{-1} \Sigma_0$$

og

$$(4.39) \quad \Sigma^{-1} = nQ_n \Sigma_0^{-1}.$$

Siden Q er diagonal og Σ_0^{-1} symmetrisk, spiller det ingen rolle om vi skriver $Q \Sigma_0^{-1}$ eller $\Sigma_0^{-1} Q$.

Vi forutsetter her og i det følgende at matrisene P_n , Q_n og Z er ikke-stokastiske. Alle resultater vil gjelde også dersom dette ikke er tilfelle, bortsett fra at vi i en del uttrykk må skrive plim istedenfor lim.

4.3 Estimering ved to-trinns generaliserte minste kvadraters metode (2SGLS)

Dersom matrisen Σ er kjent, er den beste lineære estimatoren av vektoren β i (4.20) gitt ved

$$(4.40) \quad \underline{b}^0 = (Z'\Sigma^{-1}Z)^{-1}Z'\Sigma^{-1}\underline{y}.$$

Dette er den såkalte generaliserte minste kvadraters estimator (GLS- "generalized least squares"), eller Aitken-estimatoren.

GLS-estimatoren \underline{b}^0 er "best" (optimal) i følgende forstand. Den er forventningsrett, og kovariansmatrisen til \underline{b}^0 er gitt ved

$$(4.41) \quad \Sigma_{b^0} = (Z'\Sigma^{-1}Z)^{-1}.$$

Dette er den "minste" kovariansmatrise noen lineær forventningsrett estimator av β kan ha. Med "lineær" mener vi her "lineær i \underline{y} ". Estimatoren \underline{b}^0 er BLU ("best linear unbiased") (jfr. avsnitt A.3.1.).

I de fleste praktiske tilfelle er matrisen Σ ikke kjent. \underline{b}^0 er da strengt tatt ikke en estimator, idet den avhenger av ukjente parametre. Dersom vi imidlertid kan finne en konsistent estimator av Σ , la oss kalle denne $\hat{\Sigma}$, kan vi bruke følgende vektor som estimator for β :

$$(4.42) \quad \underline{b} = (Z'\hat{\Sigma}^{-1}Z)^{-1}Z'\hat{\Sigma}^{-1}\underline{y}.$$

I engelsk-språklig litteratur går slike estimatører under navnet "feasible Aitken estimators" (feasible=gjennomførbar), se f.eks. Dhrymes (1970).

Hvordan finner vi nå en konsistent estimator av Σ ? Det finnes flere utveier. Den enkleste består i innsetting av y_{gj} for p_{gj} i uttrykkene (4.27) og (4.28). Istedenfor de ukjente parametrene p_{gj} bruker vi altså de observerte andelene y_{gj} . Denne metoden vil vi kalle to-trinns generalisert minste kvadraters metode (2SGLS - "2-stage generalized least squares"). Første trinn består i å danne den konsistente estimator $\hat{\Sigma}$ av Σ , og annet trinn i å regne ut uttrykket (4.42).

Vi forutsetter her at alle andelene y_{gj} er ekte større enn 0 og mindre enn 1, dersom dette ikke er tilfelle blir matrisen $\hat{\Sigma}$ singular. I praksis er denne forutsetningen ikke alltid oppfylt - i avsnitt 8.3 kommer vi tilbake til hvordan dette problemet kan behandles. Det er åpenbart at 2SGLS bare kan fungere brukbart dersom antallet observasjoner (n_g) innen hver gruppe g er rimelig stort. Dersom f.eks. noen av de uavhengige variable er kontinuerlige, vil dette kravet ikke være oppfylt, jfr. avsnitt A.2.

Dhrymes (1970), s. 161-166 viser at 2SGLS-estimatøren er asymptotisk ekvivalent med GLS, altså med den beste lineære forventningsrette estimator vi kan finne. Begge estimatører har asymptotisk fordeling gitt ved

$$(4.43) \quad \sqrt{n}(\underline{b}-\beta) \xrightarrow{L} N(0, \lim_{n \rightarrow \infty} n(Z'\Sigma^{-1}Z)^{-1}).$$

Dersom vi har bare grupperte data og matrisen Z bare består av dummy-variable for de ulike kategorier, vil Z være konstant og ha samme dimensjon uansett utvalgets størrelse. Ethvert nytt individ i som tilføres utvalget vil komme til uttrykk ikke som en ny rad i matrisen X (og $r-1$ nye rader i Z), men ved at n_g og n_{gjj} øker med 1, dersom individ i tilhører gruppe g og har verdi j på den avhengige variable. Kovariansmatrisen i den asymptotiske fordelingen kan da skrives

$$(4.44) \quad \lim_{n \rightarrow \infty} n(Z'\Sigma^{-1}Z)^{-1} = (Z'Q\Sigma_0^{-1}Z)^{-1},$$

jfr. (4.37) og (4.39).

En konsistent estimator for denne kovariansmatrisen er gitt ved

$$(4.45) \quad n\hat{\Sigma}_b = n(Z'\hat{\Sigma}^{-1}Z)^{-1} \xrightarrow{P} (Z'Q\Sigma^{-1}Z)^{-1}$$

Anta at vi er interessert i å teste hypotesen

$$(4.46) \quad H_0 : K\underline{\beta} = \underline{\kappa}$$

mot alternativet

$$(4.47) \quad H_1 : K\underline{\beta} \neq \underline{\kappa}$$

K er her en $(v \times m)$ konstant matrise av rang v og $\underline{\kappa}$ en $(v \times 1)$ konstant vektor. Dersom f.eks. $v = 1$, $K = (1 \ -1 \ 0 \dots 0)'$ og $\underline{\kappa} = 0$, så er H_0 hypotesen $\beta_{11} = \beta_{12}$. Generelt betegner imidlertid H_0 et sett av v lineært uavhengige simultane hypoteser om elementene i $\underline{\beta}$.

Under H_0 gjelder, ifølge (4.43) og (A.15),

$$(4.48) \quad \sqrt{n} (K\underline{b} - \underline{\kappa}) \xrightarrow{L} N(0, \lim_{n \rightarrow \infty} nK(Z'\Sigma^{-1}Z)^{-1}K')$$

og ifølge setning A.6 og (4.45)

$$(4.49) \quad (K\underline{b} - \underline{\kappa})' [K\hat{\Sigma}_b K']^{-1} (K\underline{b} - \underline{\kappa}) \xrightarrow{L} \chi^2(v).$$

En asymptotisk test av H_0 på nivå α er dermed: forkast H_0 dersom observatoren (4.49) er større enn $(1-\alpha)$ - fraktilen i kji-kvadratfordelingen med v frihetsgrader.

En "feasible Aitken-estimator" av $\underline{\beta}$ kan som nevnt også tenkes framkommet på andre måter, f.eks. slik: (1) Sett (f.eks.) $p_{gj} = \frac{1}{r}$ ($j=1,2,\dots,r-1$) i (4.27) og (4.28) og regn ut estimatoren \underline{b}^0 , ifølge formel (4.40). Regn ut de føyede andeler $\hat{y}^0 = Z\underline{b}^0$. Gitt modellen (4.20), er \hat{y}^0 en konsistent estimator av $(p_{11}, \dots, p_{s,r-1})'$. (2) Bruk disse estimatene til å danne matrisen $\hat{\Sigma}^0$ og regn ut

$$(4.50) \quad \underline{b}^1 = (Z'(\hat{\Sigma}^0)^{-1}Z)^{-1}Z'(\hat{\Sigma}^0)^{-1}y$$

og eventuelt

$$(4.51) \quad \hat{y}^1 = Z\underline{b}^1.$$

(3) Fortsett eventuelt ved å danne $\hat{\Sigma}^1$ på grunnlag av \hat{y}^1 , og så videre som en iterasjonsprosess i t trinn, inntil avviket mellom \hat{y}^t og \hat{y}^{t-1} er blitt tilstrekkelig lite.

Denne metoden kunne vi passende kalle iterativ minste kvadraters metode (IGLS). McGillivray (1970) har vist at IGLS er konsistent. IGLS kan anvendes også på individdata, dvs. når $n_g = 1$. Til gjengjeld er den atskillig mer komplisert rent regneteknisk. Anvendt på den lineære modell har metoden dessuten én alvorlig svakhet, nemlig at $\hat{y}^0, \hat{y}^1, \dots$ kan komme til å inneholde negative elementer. Dette vil gi enkelte negative varianser på diagonalen i $\hat{\Sigma}^0, \hat{\Sigma}^1, \dots$, og tilsvarende positive kovarianser utenfor diagonalen. Vi anbefaler derfor generelt ikke å bruke denne metoden til estimering av den lineære modell.

Matrisen Σ_1 eller $\hat{\Sigma}$, kan invertteres analytisk, se f.eks. Lee et al. (1970), kap. 6.3. Det vises der at den inverse av Σ er av en særdeles enkel form, nemlig

$$(4.52) \quad \Sigma^{-1} = \begin{bmatrix} \Sigma^{11} & \Sigma^{12} & \dots & \Sigma^{1,r-1} \\ \Sigma^{12} & \Sigma^{22} & \dots & \Sigma^{2,r-1} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma^{1,r-1} & \Sigma^{2,r-1} & \dots & \Sigma^{r-1,r-1} \end{bmatrix}$$

der

$$(4.53) \quad \Sigma^{jj} = \begin{bmatrix} n_1(p_{1r}^{-1} + p_{1j}^{-1}) & 0 & \dots & 0 \\ 0 & n_2(p_{2r}^{-1} + p_{2j}^{-1}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & n_s(p_{sr}^{-1} + p_{sj}^{-1}) \end{bmatrix} \quad (j=1,2,\dots,r-1)$$

og

$$(4.54) \quad \Sigma^{jk} = \begin{bmatrix} n_1 p_{1r}^{-1} & 0 & \dots & 0 \\ 0 & n_2 p_{2r}^{-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & n_s p_{sr}^{-1} \end{bmatrix} \quad (j < k).$$

Også den inverse Σ^{-1} består altså av $(r-1)^2$ diagonale delmatriser.

Ved å sette de konsistente estimatorer for p_{gj} , y_{gj} , direkte inn i uttrykket for Σ^{-1} , finner vi $\hat{\Sigma}^{-1}$ slik denne matrisen inngår i (4.42). Derved slipper vi å invertere $\hat{\Sigma}$ numerisk, noe som er svært arbeidsbesparende dersom r og/eller s er store.

Vi har skissert en metode til estimering av parametrene β i den lineære modellen (4.20). Vår egentlige interesse retter seg ofte mot sannsynlighetene p_{gj} snarere enn mot β . Estimer av p_{gj} får vi imidlertid som de "føyede verdier" i modellen (4.20):

$$(4.55) \quad \hat{p} = Z\hat{\beta} = Z(Z'\hat{\Sigma}^{-1}Z)^{-1}Z'\hat{\Sigma}^{-1}y.$$

Formel (4.55) definerer estimer \hat{p}_{gj} tilsvarende sannsynlighetene p_{gj} for $g = 1, 2, \dots, s$; $j = 1, 2, \dots, r-1$. Estimer for sannsynlighetene tilhørende r -te utfall følger naturlig av betingelsen (4.10a), slik at

$$(4.56) \quad \hat{p}_{gr} = 1 - \sum_{j=1}^{r-1} \hat{p}_{gj} \quad (g=1, 2, \dots, s).$$

Det spiller ingen rolle for estimatene \hat{p}_{gj} ($g=1, 2, \dots, r$; $j=1, 2, \dots, r$) hvilket av de r utfall som velges om "basis".

4.4 Sammenhengen mellom individdata og grupperte data

Istedenfor å komprimere datamaterialet til r s grupper og danne regresjonen av de $(r-1)$ s andelene y med hensyn på Z , kunne en tenke seg å nytte hvert individ som en observasjon. Vi ville da få ialt $(r-1) \sum_{g=1}^s n_g$ observasjoner. Hvert individ er da karakterisert gjennom et sett av $r-1$ "avhengige" binærevariable $(y_{i1}^* \dots y_{i,r-1}^*)$ og av en vektor x_i^* av "uavhengige" variable, slik at $x_i^* = x_g^*$ dersom individ i er medlem av gruppe g . Likningssystemet (4.20) kan da skrives

$$(4.57) \quad \underline{y}^* = Z^* \underline{\beta} + \underline{u}^*$$

der

$$(4.58) \quad \underline{y}^* = (y_{11}^* \dots y_{n1}^* \quad y_{12}^* \dots y_{n2}^* \quad \dots \quad y_{1,r-1}^* \dots y_{n,r-1}^*)'$$

$$(4.59) \quad \underline{u}^* = (u_{11}^* \dots u_{n1}^* \quad u_{12}^* \dots u_{n2}^* \quad \dots \quad u_{1,r-1}^* \dots u_{n,r-1}^*)'$$

$$(4.60) \quad Z^* = I_{r-1} \otimes X^*$$

og

$$(4.61) \quad X^* = \begin{bmatrix} x_1^* \\ x_2^* \\ \vdots \\ x_n^* \end{bmatrix} = \begin{bmatrix} i_{n_1} \otimes x_1^* \\ \vdots \\ i_{n_s} \otimes x_s^* \end{bmatrix}.$$

y_{ij}^* er 1 dersom individ i har utfall j , ellers 0, jfr. (4.7). n er det totale antall individer i observasjonsmaterialet, dvs.

$$(4.62) \quad n = \sum_{g=1}^s n_g.$$

Dessuten gjelder sammenhengen

$$(4.63) \quad y_{gj} = \frac{1}{n_g} \sum_{i=q(g)+1}^{q(g+1)} y_{ij}^*$$

der

$$(4.64) \quad q(g) = \sum_{h=1}^{g-1} n_h \quad (g=1, 2, \dots, s+1).$$

Modellen (4.57) oppstår som et spesialtilfelle av (4.20) dersom alle n_g er lik 1, dvs. det er bare ett individ i hver "gruppe".

Fordelingsegenskapene til restleddene u_{ij}^* blir som for u_{gj} , gitt ved (4.12) - (4.15), med 1 innsett for alle n_g .

4.5 Estimering ved vanlig minste kvadraters metode (OLS)

Vanlig minste kvadraters estimatoren (OLS - "ordinary least squares") for modellen (4.57) er gitt ved

$$(4.65) \quad \hat{\beta} = (Z^{*'} Z^*)^{-1} Z^{*'} y^*.$$

Denne estimatoren $\hat{\beta}$ kan også skrives

$$(4.66) \quad \hat{\beta} = (Z' MZ)^{-1} Z' My,$$

der matrisen M er definert ved (4.31). Det er nemlig lett å vise at

$$(4.67) \quad Z^{*'} Z^* = Z' MZ$$

og

$$(4.68) \quad Z^{*'} y^* = Z' My.$$

Estimatoren $\hat{\beta}$ er forventningsrett. Kovariansmatrisen til $\hat{\beta}$ er gitt ved

$$(4.69) \quad \Sigma_{\hat{\beta}} = (Z' MZ)^{-1} Z' M \Sigma MZ (Z' MZ)^{-1}.$$

Ved å sette inn (4.22), (4.26) og (4.30) i (4.69), og bruke regnereglene for Kronecker-produkt, er det lett å vise at matrisen $\Sigma_{\hat{\beta}}$ består av $(r-1)^2$ blokker, der blokk (j,k) er $(m \times m)$ - matrisen

$$(4.70) \quad E[(\hat{\beta}_j - \beta_j)(\hat{\beta}_k - \beta_k)'] = (X'NX)^{-1} X'N \Sigma_{jk} NX(X'NX)^{-1} \quad (j \leq k).$$

OLS-estimatoren $\hat{\beta}$ har asymptotisk fordeling gitt ved

$$(4.71) \quad \sqrt{n} (\hat{\beta} - \beta) \xrightarrow{L} N(0, \lim_{n \rightarrow \infty} n \Sigma_{\hat{\beta}}).$$

Dersom vi har bare grupperte data og Z er konstant, har vi analogt med (4.44),

$$(4.72) \quad \lim_{n \rightarrow \infty} n \Sigma_{\hat{\beta}} = (Z' QZ)^{-1} Z' Q \Sigma_0 Z (Z' QZ)^{-1}.$$

Resultatet følger av (4.69), (4.38) og (4.37).

En konsistent estimator av $\Sigma_{\hat{\beta}}$, $\hat{\Sigma}_{\hat{\beta}}$, får vi ved innsetting av y_{gj} for p_{gj} i uttrykket for $\Sigma_{\hat{\beta}}$ (4.69). Analogt med hva vi så i 2SGLS-tilfellet, kan vi asymptotisk teste enhver hypotese av formen (4.46) ved å observere at, under H_0 ,

$$(4.73) \quad (K\hat{\beta} - \zeta)' [K\hat{\Sigma}_{\hat{\beta}}K']^{-1} (K\hat{\beta} - \zeta) \xrightarrow{L} \chi^2(\nu).$$

Estimering ved hjelp av formelen (4.66) er en form for veiet regresjon anvendt på modellen (4.20). Siden M er en diagonal matrise, innebærer formel (4.66) at vi veier observasjonene i hver gruppe g med kvadratroten av gruppestørrelsen n_g . Dette er intuitivt rimelig siden standardavviket på hvert gruppegjennomsnitt y_{gj} er omvendt proporsjonalt nettopp med $\sqrt{n_g}$, jfr. (4.14). Denne form for veiet regresjon anvendt på de grupperte data gir altså samme resultat ($\hat{\beta}$) som vanlig (uveiet) minste kvadraters metode anvendt på individdata. Det er derfor naturlig i vårt tilfelle å kalle estimatoren $\hat{\beta}$, gitt ved (4.66), for OLS-estimatoren av β .

Definer den diagonale matrisen $M^{\frac{1}{2}}$ slik at

$$(4.74) \quad M^{\frac{1}{2}} M^{\frac{1}{2}} = M,$$

dvs. diagonalelementene i $M^{\frac{1}{2}}$ består av kvadratrøttene $\sqrt{n_1}$, $\sqrt{n_2}$, osv. Estimatoren $\hat{\beta}$ er da ekvivalent med OLS anvendt på den transformerte modellen

$$(4.75) \quad M^{\frac{1}{2}} \underline{y} = M^{\frac{1}{2}} Z \underline{\beta} + M^{\frac{1}{2}} \underline{u}.$$

Definer matrisen S slik at

$$(4.76) \quad S'S = \hat{\Sigma}^{-1}.$$

Dersom vi innfører et krav om at den skal ha bare positive diagonalelementer, er S entydig definert som en ikke-singulær triangulær matrise (se Dhrymes (1970), s.579-580). 2SGLS-estimatoren \underline{b} (4.42) kan nå tolkes som OLS anvendt på den transformerte modell.

$$(4.77) \quad S\underline{y} = SZ\underline{\beta} + S\underline{u}.$$

Begge estimatorene $\hat{\beta}$ og \underline{b} får vi altså ved å utføre uveiet regresjon på en bestemt transformasjon av de grupperte data \underline{y} og X . Begge transformasjonene innebærer multiplikasjon med kvadratroten av gruppestørrelsen n_g , 2SGLS er imidlertid vesensforskjellig fra OLS fordi (i) 2SGLS tar hensyn til at restleddsvariansen $\text{var}(u_{gj})$ avhenger ikke bare av n_g , men også av $p_{gj}(1-p_{gj})$, og (ii) 2SGLS tar hensyn til avhengigheten (kovariansen) mellom de ulike mulige utfall for samme gruppe individer.

Punkt (ii) innebærer at dataene (y_{gj}) for utfall j bidrar informasjon ikke bare til estimering av delvektoren β_j , men også til estimering av β_k ($k \neq j$). (Husk at $\underline{\beta} = (\beta_1' \beta_2' \dots \beta_{r-1}')'$.) Dette er ikke tilfelle med OLS-estimatoren (4.66), og vi kan her faktisk estimere hver delvektor separat:

$$(4.78) \quad \hat{\beta}_j = (X'NX)^{-1} X'N\underline{y}_{.j} \quad (j=1,2,\dots,r).$$

(4.66) må gi samme resultat $\hat{\underline{\beta}} = (\hat{\beta}_1' \hat{\beta}_2' \dots \hat{\beta}_{r-1}')'$ som (4.78) fordi begge matrisene Z og M som inngår i uttrykket (4.66), er blokkdiagonale. (4.66) og (4.78) er derfor algebraisk identiske.

Ved innsetting av estimatene $\hat{\beta}_j$ for de ukjente parametre β_j finner vi de føyede verdiene $\hat{y}_{.j}$:

$$(4.79) \quad \hat{y}_{.j} = X\hat{\beta}_j \quad (j=1,2,\dots,r).$$

Et rimelig krav til estimeringsmetoden er at

$$(4.80) \quad \sum_{j=1}^r \hat{y}_{.j} = \underline{j},$$

altså at de estimerte sannsynlighetene \hat{y}_{gj} , summert over alle mulige utfall j , blir lik 1, til tross for at de blir estimert separat.

Det er lett å vise at OLS faktisk oppfyller dette kravet, forutsatt at hver delvektor β_j inneholder et konstantledd, f.eks. ved at første kolonne i matrisen X består av bare ett-tall (se f.eks. H.T. Amundsen (1976)).

I avsnitt 4.3 konstaterte vi at vi ved 2SGLS-estimering fritt kan velge ett av de r mulige utfall som "basis", uten at dette påvirker resultatene \hat{p}_{gj} ($g=1,2,\dots,s; j=1,2,\dots,r$). Relasjon (4.80) impliserer at det samme gjelder for OLS-estimering.

(4.80) innebærer imidlertid ikke at

$$(4.81) \quad 0 \leq \hat{y}_{gj} \leq 1$$

for alle g og j . Vi kan f.eks. ha $\hat{y}_{g1} = 1.2$, $\hat{y}_{g2} = 0.1$, $\hat{y}_{g3} = -0.3$ ($r=3$). Dette gjelder for OLS-såvel som for 2SGLS-estimering.

4.6 Estimering ved sannsynlighetsmaksimering (ML)

Vi skal her kort nevne en tredje metode som i prinsippet kan nyttes til estimering av modellen (4.20), nemlig sannsynlighetsmaksimering (ML). La n_{gj} være antall individer i gruppe g som har utfall nr. j , dvs.

$$(4.82) \quad n_{gj} = n_g y_{gj}$$

Hver vektor $n_g = (n_{g1}, \dots, n_{gr})'$ er da som nevnt multinomisk fordelt med parametre p_{g1}, \dots, p_{gr} og n_g . "Log-likelihood" -funksjonen til hele observasjonsmaterialet $[n_1, n_2, \dots, n_s]$ er gitt ved

$$(4.83) \quad \log L = \sum_{g=1}^s \left\{ \log(n_g!) + \sum_{j=1}^r [n_{gj} \log p_{gj} - \log(n_{gj}!)] \right\}.$$

ML-estimatorene finnes ved å maksimere denne funksjonen med hensyn på parametrene β , idet vi setter inn for

$$(4.84) \quad \left\{ \begin{array}{l} p_{gj} = x_{g-j}' \beta_j \quad (j=1, 2, \dots, r-1) \\ p_{gr} = 1 - \sum_{j=1}^{r-1} p_{gj} = 1 - \sum_{j=1}^{r-1} x_{g-j}' \beta_j \end{array} \right\}.$$

jfr. (4.10a), (4.10b) og (4.11).

Innsetting av (4.84) i (4.83) gir

$$(4.85) \quad \log L = \sum_{g=1}^s \left\{ \sum_{j=1}^{r-1} n_{gj} \log(x_{g-j}' \beta_j) - n_{gr} \log\left(1 - \sum_{j=1}^{r-1} x_{g-j}' \beta_j\right) \right\} + \text{konstantledd.}$$

Ved å derivere denne funksjonen partielt med hensyn på alle β_{jk} og sette de deriverte lik null, får vi et (ikke-lineært) system av $(r-1)m$ likninger

$$(4.86) \quad \frac{\partial \log L}{\partial \beta_{jk}} = \sum_{g=1}^s \left[n_{gj} x_{gk} / (x_{g-j}' \beta_j) - n_{gr} x_{gk} / \left(1 - \sum_{j=1}^{r-1} x_{g-j}' \beta_j\right) \right] = 0 \quad (j=1, 2, \dots, r-1; k=1, 2, \dots, m).$$

Løsningen av dette likningssystemet er ML-estimatene av β .

Ved å betrakte (4.83) innser en lett en fordel som ML-estimering har framfor OLS eller 2SGLS: Siden $\log L$ avhenger av p_{gj} bare gjennom den logaritmiske funksjon, som ikke er definert for negative argument, er vi garantert mot å få negative sannsynlighetsestimater \hat{p}_{gj} for alle de kombinasjoner av x_g -vektorer og mulige utfall j som opptrer i utvalget. For andre kombinasjoner av uavhengige variable enn de som opptrer i utvalget, gir imidlertid metoden ingen slik garanti.

Baksiden av denne medaljen er at en enkelt slenger ("outlier") i utvalget kan være nok til å tvinge parameterestimaterne drastisk nedover. Metoden er m.a.o. langt mindre robust overfor ekstreme eller "ville" observasjoner enn f.eks. OLS og 2SGLS.

I binær-tilfellet $r=2$ kan vi skrive (4.85) på noe forenklet form:

$$(4.87) \quad \log L = \sum_{g=1}^s n_g [y_{g1} \cdot \log(x'_{g1} \beta_1) - (1-y_{g1}) \log(1-x'_{g1} \beta_1)] + \text{konstantledd.}$$

Derivasjon gir her

$$(4.88) \quad \frac{\partial \log L}{\partial \beta_{1k}} = \sum_{g=1}^s n_g \left[\frac{y_{g1} x_{gk}}{x'_{g1} \beta_1} - \frac{(1-y_{g1}) x_{gk}}{1-x'_{g1} \beta_1} \right] = 0 \quad (k=1,2,\dots,m).$$

Minimering av (4.85) eller (4.87) kan gjennomføres ved hjelp av generelle algoritmer for løsning av ikke-lineære likningssystem.

4.7. Estimering ved minimum Neyman kji-kvadrat ($\min x_N^2$)

En svært vanlig metode til estimering av parametrene i en kontingenstabell er minimum kji-kvadrat-metoden. Anvendt på vår modell vil denne metoden bestå i minimering av uttrykket

$$(4.89) \quad x_p^2 = \sum_{g=1}^s n_g \sum_{j=1}^r [(y_{gj} - p_{gj})^2 / p_{gj}]$$

med hensyn på parametrene β . Sannsynlighetene p_{gj} er selvsagt funksjoner av β , jfr. (4.84). Observatoren x_p^2 er kjent som Pearson kji-kvadrat-observatoren. Innsatt for minimum kji-kvadrat-estimaterne av β er x_p^2 asymptotisk kji-kvadrat fordelt med v frihetsgrader, dersom (4.84) er den sanne modell. Her er v lik forskjellen mellom antall uavhengige celler i tabellen og antall parametre som estimeres, i vårt tilfelle:

$$(4.90) \quad v = (r-1)(s-m).$$

Minimering av x_p^2 byr på endel numeriske problemer bl.a. fordi parametrene også forekommer i nevner. Neyman (1949) har foreslått en alternativ observator, nemlig

$$(4.91) \quad x_N^2 = \sum_{g=1}^s n_g \sum_{j=1}^r [(y_{gj} - p_{gj})^2 / y_{gj}].$$

Her er altså sannsynlighetene p_{gj} i nevner erstattet med de observerte andelene y_{gj} . Denne observatoren kalles modifisert Pearson kji-kvadrat eller Neyman kji-kvadrat. De såkalte minimum Neyman kji-kvadrat-estimatorene oppnås ved minimering av x_N^2 m.h.p. β . Innsatt for disse estimatorene er x_N^2 , i likhet med x_p^2 , asymptotisk kji-kvadrat fordelt med v frihetsgrader, gitt (4.11).

I binær-tilfellet ($r=2$) kan x_N^2 observatoren skrives på formen

$$(4.92) \quad x_N^2 = \sum_{g=1}^s n_g \frac{(y_{g1} - p_{g1})^2}{y_{g1}(1-y_{g1})}.$$

Vi kan nå vise

Setning 4.1. Minimering av Neyman kji-kvadrat-observatoren (4.91) er algebraisk ekvivalent med 2SGLS (4.42).

Bevis: Definer $\underline{y}_g = (y_{g1}, \dots, y_{g,r-1})'$, $\underline{p}_g = (p_{g1}, \dots, p_{g,r-1})'$, $\underline{u}_g = (u_{g1}, \dots, u_{g,r-1})'$,

$$\Sigma_g = \text{Cov}(\underline{u}_g), \hat{\Sigma}_g = \Sigma_g \Big|_{p_{gj}=y_{gj}}$$

Systemet

$$(4.93) \quad \underline{y}_g = \underline{p}_g + \underline{u}_g = [I_{r-1} \otimes \underline{x}_g'] \underline{\beta} + \underline{u}_g \quad (g=1, \dots, s)$$

er identisk med modellen (4.20), bortsett fra at observasjonene y_{gj} nå er hovedsortert etter første indeks g , mot tidligere etter j .

2SGLS består i minimering av uttrykket

$$(4.94) \quad \xi(\underline{\beta}) = (\underline{y} - \underline{p})' \hat{\Sigma}^{-1} (\underline{y} - \underline{p}) = (\underline{y} - \underline{Z}\underline{\beta})' \hat{\Sigma}^{-1} (\underline{y} - \underline{Z}\underline{\beta})$$

som, ifølge (4.93), er ekvivalent med

$$(4.95) \quad \xi(\underline{\beta}) = \sum_{g=1}^s (\underline{y}_g - \underline{p}_g)' \hat{\Sigma}_g^{-1} (\underline{y}_g - \underline{p}_g)$$

Nå er

$$(4.96) \quad \Sigma_g = \frac{1}{n_g} (D_g - \underline{p}_g \underline{p}_g')$$

og

$$(4.97) \quad \Sigma_g^{-1} = n_g (D_g^{-1} + \underline{p}_{gr} \underline{p}_{gr}')^{-1}$$

der D_g er en diagonal matrise med $(p_{g1}, \dots, p_{g,r-1})$ på diagonalen. For $\hat{D}_g = D_g \Big|_{p_{gj}=y_{gj}}$ har vi da

$$(4.98) \quad \begin{aligned} \xi(\underline{\beta}) &= \sum_{g=1}^s n_g (\underline{y}_g - \underline{p}_g)' \hat{D}_g^{-1} (\underline{y}_g - \underline{p}_g) + \sum_{g=1}^s n_g y_{gr}^{-1} (\underline{y}_g - \underline{p}_g)' \underline{1} \underline{1}' (\underline{y}_g - \underline{p}_g) \\ &= \sum_{g=1}^s n_g \sum_{j=1}^{r-1} [(y_{gj} - p_{gj})^2 / y_{gj}] + \sum_{g=1}^s n_g y_{gr}^{-1} \left[\sum_{j=1}^{r-1} (y_{gj} - p_{gj}) \right]^2 \\ &= \sum_{g=1}^s n_g \sum_{j=1}^r [(y_{gj} - p_{gj})^2 / y_{gj}] = \chi_N^2, \text{ q.e.d.} \end{aligned}$$

Det nest siste likhetstegnet følger av $y_{gr} = 1 - \sum_{j=1}^{r-1} y_{gj}$ og $p_{gr} = 1 - \sum_{j=1}^{r-1} p_{gj}$. Beviset bygger i stor grad på Clickner (1976), som viser det samme i spesialtilfellet $s=1$.

Ifølge dette resultatet har altså 2SGLS-estimatorene ikke bare samme sannsynlighetsfordeling som minimum Neyman kji-kvadrat-observatorene, de to sett av estimatører er endatil algebraisk identiske. 2SGLS-estimatorene har dermed også alle min χ_N^2 -estimatorenes egenskaper. Om min χ_N^2 -estimatorene vet vi at de er såkalt BAN ("best asymptotically normal"), dvs. de er konsistente og konvergerer mot en multi-normal fordeling med den "minste" kovariansmatrise noen konsistent estimator kan ha, ifølge Cramér-Raos lov (Neyman (1949)). Dette gjelder m.a.o. også 2SGLS.

Vi kan videre vise

Setning 4.2. La \underline{b}_0 betegne 2SGLS-estimatoren av $\underline{\beta}$ beregnet under bibetingelsen

$$(4.46) \quad H_0: K\underline{\beta} = \underline{\kappa},$$

der K er av dimensjon $(v \times m)$ og rang v . La \underline{b} som før betegne 2SGLS-estimatoren (4.42) beregnet uten bibetingelser. La endelig $\xi(\underline{b}_0)$ og $\xi(\underline{b})$ betegne Neyman kji-kvadrat-observatorene tilsvarende de to estimatorene, jfr. (4.94) - (4.98).

Differansen mellom $\xi(\underline{b}_0)$ og $\xi(\underline{b})$ er da

$$(4.99) \quad \xi(\underline{b}_0) - \xi(\underline{b}) = (K\underline{b} - \underline{\kappa})' [K(Z'\hat{\Sigma}^{-1}Z)^{-1}K']^{-1} (K\underline{b} - \underline{\kappa}),$$

dvs. identisk med testobservatoren (4.49), som ble utledet i avsnitt 4.3.

Bevis: Definer Lagrange-funksjonen

$$(4.100) \quad \begin{aligned} \omega(\underline{\beta}) &= \xi(\underline{\beta}) + \underline{\pi}'(K\underline{\beta} - \underline{\kappa}) \\ &= (\underline{y} - Z\underline{\beta})'\hat{\Sigma}^{-1}(\underline{y} - Z\underline{\beta}) + \underline{\pi}'(K\underline{\beta} - \underline{\kappa}), \end{aligned}$$

der $\underline{\pi}$ er en vektor av Lagrange-multiplikatorer. \underline{b}_0 finnes ved minimering av $\omega(\underline{\beta})$ m.h.p. $\underline{\beta}$ og $\underline{\pi}$. Ved å bruke reglene for derivasjon av matriseuttrykk (se f.eks. Theil (1971), s. 31-32), finner vi førsteordensbetingelsene

$$(4.101) \quad \frac{\partial \omega}{\partial \underline{\beta}} = 2Z'\hat{\Sigma}^{-1}Z\underline{\beta} - 2Z'\hat{\Sigma}^{-1}\underline{y} + K'\underline{\pi} = \underline{0}$$

$$(4.102) \quad \frac{\partial \omega}{\partial \underline{\pi}} = K\underline{\beta} - \underline{\kappa} = \underline{0}.$$

Noter først at \underline{b} finnes som løsningen på systemet (4.101) i spesialtilfellet $\underline{\pi} = \underline{0}$ (ingen bibetingelser). Dette impliserer

$$(4.103) \quad Z'\hat{\Sigma}^{-1}(\underline{y} - Z\underline{b}) = \underline{0}.$$

Vi multipliserer nå (4.101) med $K(Z'\hat{\Sigma}^{-1}Z)^{-1}$, setter inn $K\underline{\beta} = \underline{\kappa}$ (4.102), noterer at matrisen $K(Z'\hat{\Sigma}^{-1}Z)^{-1}K'$ er ikke-singulær siden K har full rang, og finner

$$(4.104) \quad \begin{aligned} \underline{\pi} &= 2 [K(Z'\hat{\Sigma}^{-1}Z)^{-1}K']^{-1} [K(Z'\hat{\Sigma}^{-1}Z)^{-1}Z'\hat{\Sigma}^{-1}\underline{y} - \underline{\kappa}] \\ &= 2 [K(Z'\hat{\Sigma}^{-1}Z)^{-1}K']^{-1} (K\underline{b} - \underline{\kappa}). \end{aligned}$$

Ved å sette dette uttrykket inn i (4.101) og bruke (4.42), finner vi at 2SGLS-løsningen under H_0 kan skrives

$$(4.105) \quad \underline{b}_0 = \underline{b} - (Z'\hat{\Sigma}^{-1}Z)^{-1}K'[K(Z'\hat{\Sigma}^{-1}Z)^{-1}K']^{-1}(K\underline{b} - \underline{\kappa}).$$

Vi har dessuten

$$(4.106) \quad \begin{aligned} \xi(\underline{b}_0) &= (\underline{y} - Z\underline{b}_0)'\hat{\Sigma}^{-1}(\underline{y} - Z\underline{b}_0) \\ &= [\underline{y} - Z\underline{b} - Z(\underline{b}_0 - \underline{b})]'\hat{\Sigma}^{-1}[\underline{y} - Z\underline{b} - Z(\underline{b}_0 - \underline{b})] \\ &= (\underline{y} - Z\underline{b})'\hat{\Sigma}^{-1}(\underline{y} - Z\underline{b}) - 2(\underline{b}_0 - \underline{b})'Z'\hat{\Sigma}^{-1}(\underline{y} - Z\underline{b}) + (\underline{b}_0 - \underline{b})'Z'\hat{\Sigma}^{-1}Z(\underline{b}_0 - \underline{b}). \end{aligned}$$

Her er det første leddet lik $\varepsilon(\underline{b})$ og det andre leddet lik 0 ifølge (4.103). Ved å kombinere med (4.105) finner vi således

$$\begin{aligned} (4.107) \quad & \varepsilon(\underline{b}_0) - \varepsilon(\underline{b}) \\ &= (\underline{Kb} - \underline{\kappa})' [K(Z'\hat{\Sigma}^{-1}Z)^{-1}K']^{-1} K(Z'\hat{\Sigma}^{-1}Z)^{-1} Z'\hat{\Sigma}^{-1} Z(Z'\hat{\Sigma}^{-1}Z)^{-1} K' [K(Z'\hat{\Sigma}^{-1}Z)^{-1}K']^{-1} (\underline{Kb} - \underline{\kappa}) \\ &= (\underline{Kb} - \underline{\kappa})' [K(Z'\hat{\Sigma}^{-1}Z)^{-1}K']^{-1} (\underline{Kb} - \underline{\kappa}) \quad \text{q.e.d.} \end{aligned}$$

Setning 4.2 gir oss en snarvei til beregning av testobservatoren (4.49). Istedenfor å regne ut dette uttrykket slik det står, kan vi ganske enkelt ta differensen mellom (Neyman) kji-kvadrat-observatorene under H_0 og H_1 . Forutsatt at vi tester mot et alternativ H_1 av formen (4.47), er denne differansen asymptotisk kji-kvadratfordelt under H_0 med ν frihetsgrader.

I kapittel 9 kommer vi tilbake til hvordan setning 4.2 kan brukes til å avgjøre hvilke samspills-effekter som bør være med i sannsynlighetsmodellen.

4.8 Sammenlikning mellom OLS og 2SGLS

Vi konkluderte i forrige avsnitt at 2SGLS er asymptotisk optimal. Dhrymes (1970), s. 159-160 viser at 2SGLS har asymptotisk mindre kovariansmatrise enn OLS, altså at matrisen

$$(4.108) \quad \Delta_{\Sigma} = \lim_{n \rightarrow \infty} n \Sigma_{\hat{\beta}} - \lim_{n \rightarrow \infty} n \Sigma_{\underline{b}}$$

er positiv definit, slik at av de to estimatorer er 2SGLS alene om å være asymptotisk optimal. Teorien kan imidlertid ikke si oss generelt hvor stor Δ_{Σ} er, dvs. hvor mye bedre 2SGLS er enn OLS.

I mindre (endelige) utvalg er fordelingsegenskapene til 2SGLS-estimatoren ikke kjent. Egenskapene kan imidlertid undersøkes gjennom empiriske studier eller Monte-Carlo-eksperiment. Dette kommer vi tilbake til i avsnitt 9.2. Det er rimelig å tro at 2SGLS må være bedre enn OLS også i de fleste mindre utvalg.

H.T. Amundsen (1974 og 1976) viser at i tilfellet med bare binære forklaringsvariable og mettet modell er OLS alltid optimal og identisk med såvel ML som 2SGLS. Resultatet gjelder imidlertid ikke når en, som i vårt tilfelle, ønsker å estimere en enklere sannsynlighetsmodell enn den mettede.

5. DEN LOG-LINEÆRE SANNSYNLIGHETSMODELL

5.1 Binær avhengig variabel

La oss, istedenfor (4.2), anta

$$(5.1) \quad p_g = h(\underline{x}_g' \underline{\beta}) = [1 + \exp(-\underline{x}_g' \underline{\beta})]^{-1}$$

Funksjonen (5.1) er kjent som den logistiske funksjon og er framstilt grafisk i figur 5.1. Vi ser her straks at

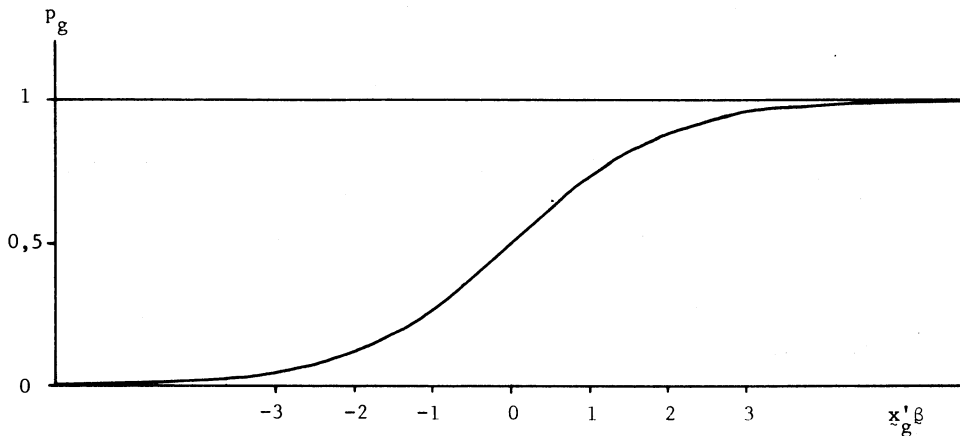
$$(5.2) \quad 0 < p_g < 1 \quad \text{for} \quad -\infty < \underline{x}_g' \underline{\beta} < \infty.$$

(5.1) kan altså tolkes som en sannsynlighet.

Ved å løse (5.1) m.h.p. $\underline{x}_g' \underline{\beta}$ finner vi

$$(5.3) \quad \log \left(\frac{p_g}{1-p_g} \right) = \underline{x}_g' \underline{\beta}.$$

Venstresiden i (5.3) kalles gjerne logit av p_g . Logaritmen til "odds"-en til p_g er lineær i parametrene $\underline{\beta}$.

Figur 5.1. Logit-funksjonen *The logit function*

Modellen (5.3) kan utledes av en mer generell modell som kalles den log-lineære modell. Dette er en modell for et kryssgruppert observasjonsmateriale, der alle grupperingsveier (eller variable) betraktes symmetrisk og vi således ikke nødvendigvis har én avhengig variabel. I vår framstilling skal vi imidlertid tenke oss at det er naturlig å betrakte en av variablene som den avhengige. I første omgang vil vi betrakte tilfellet der antall grupper for den avhengige variable er 2, og ta som eksempel en modell med i alt 3 grupperingsveier a, b og c, dvs. med én "avhengig" variabel a og to "uavhengige" variable b og c.

Den vanlige måten å skrive den log-lineære modellen på, er som følger:

$$(5.4) \quad \log p_{hjk} = \mu + \lambda_h^a + \lambda_j^b + \lambda_k^c + \lambda_{hj}^{ab} + \lambda_{hk}^{ac} + \lambda_{jk}^{bc} + \lambda_{hjk}^{abc} \quad (h, j, k=1, 2, \dots).$$

der

$$(5.5) \quad \sum_h \lambda_h^a = \sum_j \lambda_j^b = \sum_k \lambda_k^c = 0$$

$$(5.6) \quad \sum_h \lambda_{hj}^{ab} = \sum_j \lambda_{hj}^{ab} = \sum_h \lambda_{hk}^{ac} = \sum_k \lambda_{hk}^{ac} = \sum_j \lambda_{jk}^{bc} = \sum_k \lambda_{jk}^{bc} = 0$$

$$(5.7) \quad \sum_h \lambda_{hjk}^{abc} = \sum_j \lambda_{hjk}^{abc} = \sum_k \lambda_{hjk}^{abc} = 0 \quad (h, j, k=1, 2, \dots).$$

Symbolet p_{hjk} , med 3 fotskrifter, står her for den ubetingede, simultane sannsynlighet for at første variabel (a) antar verdien h, annen variabel (b) antar verdien j og tredje variabel (c) antar verdien k. Parameteren μ er definert som logaritmen til det geometriske gjennomsnitt av alle p_{hjk} , som er ekvivalent med det aritmetiske gjennomsnitt av logaritmene:

$$(5.8) \quad \mu = \frac{1}{s_a s_b s_c} \sum_{h=1}^{s_a} \sum_{j=1}^{s_b} \sum_{k=1}^{s_c} \log p_{hjk},$$

der s_a , s_b og s_c er antall grupper for de tre grupperingsveiene a, b og c, henholdsvis.

λ -parametrene med en, to og tre toppskriffter uttrykker henholdsvis en-, to- og trefaktor-effektene mellom de tre variablene. For at alle parametrene skal være identifiserbare bruker vi normeringsreglene (5.5) - (5.7), som sier at alle effekter skal bli lik null når vi summerer over hvilken som helst av grupperingsveiene. En annen måte å normere på ville være å betrakte en av gruppene i hver

grupperingsvei som basisgruppe, slik vi gjorde i kapittel 2. Dette ville imidlertid ødelegge symmetrien i notasjonen. Enten vi normerer på den ene eller annen måte, så vil p_{hjk} -ene bli de samme og få de samme estimater, men de enkelte parametrene vil ha ulik tolkning.

De enkelte λ -parametre har en tolkning som det aritmetiske gjennomsnitt av logaritmene til alle "odds-ratios", f.eks. er

$$(5.9) \quad \lambda_1^a - \lambda_2^a = \frac{1}{s_b s_c} \sum_j \sum_k \log (p_{1jk}/p_{2jk})$$

og

$$(5.10) \quad \lambda_e^a = \frac{1}{s_a s_b s_c} \sum_{h \neq e} \sum_j \sum_k \log (p_{ejk}/p_{hjk}) \quad (e=1,2,\dots,s_a).$$

Det er forøvrig flere interessante måter å tolke λ -parametrene på; litteraturen på dette området er ganske omfattende. Se f.eks. Haldorsen (1976 og 1977), Clausen (1977) eller Bishop et al. (1975).

Ved å forutsette f.eks. $\lambda_{hjk}^{abc} = 0$ for alle h, j, k får vi en modell uten tre-faktor-effekter. Dersom ingen av λ -ene settes lik null har vi den mettede modell.

Modellen (5.4) generaliseres lett til en modell med t grupperingsveier og fler-faktor-effekter av grad inntil t .

Når vi betrakter a som den avhengige variable, vil vi som oftest være mindre interessert i den ubetingede sannsynligheten p_{hjk} enn i den betingede sannsynligheten

$$(5.11) \quad P(a=1|b=j, c=k) = \frac{p_{1jk}}{\sum_h p_{hjk}}.$$

Hvis vi nå tenker oss at vi nummererer alle parene (j,k) ved hjelp av indeksen g , som løper fra 1 til $s_b \cdot s_c$, så kan vi skrive denne betingede sannsynligheten som

$$(5.12) \quad p_g = P(a=1|(b,c)=g) \quad (g=1,2,\dots,s_b \cdot s_c).$$

Da er

$$(5.13) \quad \begin{aligned} \log \left(\frac{p_g}{1-p_g} \right) &= \log p_g - \log (1-p_g) \\ &= (\mu + \lambda_1^a + \lambda_j^b + \lambda_k^c + \lambda_{1j}^{ab} + \lambda_{1k}^{ac} + \lambda_{jk}^{bc} + \lambda_{1jk}^{abc}) \\ &\quad - (\mu + \lambda_2^a + \lambda_j^b + \lambda_k^c + \lambda_{2j}^{ab} + \lambda_{2k}^{ac} + \lambda_{jk}^{bc} + \lambda_{2jk}^{abc}) \\ &= (\lambda_1^a - \lambda_2^a) + (\lambda_{1j}^{ab} - \lambda_{2j}^{ab}) + (\lambda_{1k}^{ac} - \lambda_{2k}^{ac}) + (\lambda_{1jk}^{abc} - \lambda_{2jk}^{abc}) \\ &= 2(\lambda_1^a + \lambda_{1j}^{ab} + \lambda_{1k}^{ac} + \lambda_{1jk}^{abc}). \end{aligned}$$

Det andre likhetstegnet følger av (5.4) og det fjerde av (5.5) - (5.7), idet vi har sett på spesialtilfellet $s_a = 2$ (dikotom avhengig variabel).

Definer vektoren

$$(5.14) \quad \underline{\lambda} = (\lambda_1^a \lambda_{11}^{ab} \dots \lambda_{1s_b}^{ab} \lambda_{11}^{ac} \dots \lambda_{1s_c}^{ac} \lambda_{111}^{abc} \dots \lambda_{1s_b s_c}^{abc})'$$

Ved å definere en vektor \underline{z}_g av dummyvariable, som for gitt (j,k) "plukker ut" elementene

λ_1^a , λ_{1j}^{ab} , λ_{1k}^{ac} og λ_{1jk}^{abc} fra vektoren $\underline{\lambda}$, kan vi nå skrive

$$(5.15) \quad \log \left(\frac{p_g}{1-p_g} \right) = \underline{z}_g' \underline{\lambda}.$$

Likning (5.15) viser at vektoren $\underline{\lambda}$ bare er en reparametrisering av parametrene $\underline{\beta}$ som inngår i logit-modellen (5.3). Vektoren \underline{z}_g er ikke nøyaktig lik \underline{x}_g i (5.3) fordi det er brukt ulike normeringsregler. Det er imidlertid en én-entydig sammenheng mellom $\underline{\beta}$ og $\underline{\lambda}$.

Dette forholdet kan være av stor praktisk betydning. Det innebærer nemlig at estimater beregnet innenfor en generell log-lineær modell kan oversettes til tilsvarende parameterestimater i logit-modellen. Regnemaskinprogrammer for log-lineære modeller kan brukes til å estimere logitmodellen.

Når vi bruker den generelle log-lineære modellen til å finne logit-estimater, må vi imidlertid sørge for at alle en- og fler-faktor-effekter som ikke gjelder den avhengige variable, er med i parametersettet. I vårt eksempel gjelder dette leddene λ_j^b , λ_k^c og λ_{jk}^{bc} ($j=1, \dots, s_b$; $k=1, \dots, s_c$). Estimeringmetoder for den log-lineære modell kommer vi forøvrig tilbake til i avsnitt 5.3 og 5.4.

5.2 Multinær avhengig vektor

Logit-modellen definert ved (5.1) lar seg generalisere til den såkalte multinomiske logit-modell (Theil (1969)). Denne modellen kan skrives slik:

$$(5.16) \quad p_{gj} = \frac{\exp(x_{-g}^j \beta_j)}{r + \sum_{h=1}^r \exp(x_{-g}^h \beta_h)}, \quad (g=1,2,\dots,s; j=1,2,\dots,r).$$

Uttrykk (5.16) er analogt med (4.11) i den lineære modell og gir sannsynligheten for utfall j som funksjon av verdien på de uavhengige variable for individer i gruppe g.

Vi ser straks at (5.16) oppfyller alle krav til en sannsynlighetsmodell, nemlig ved at

$$(5.17) \quad 0 < p_{gj} < 1 \quad (g=1,2,\dots,s; j=1,2,\dots,r)$$

$$(5.18) \quad \sum_{j=1}^r p_{gj} = 1 \quad (g=1,2,\dots,s).$$

Modell (5.16) inneholder r parametervektorer, men siden restriksjonen (5.18) må gjelde, inneholder systemet bare r-1 uavhengige likninger. Alle r β_j -vektorer er derfor ikke identifiserbare, og vi må også her velge en normeringsregel, f.eks. $\sum_{j=1}^r \beta_j = 0$, eller $\beta_r = 0$. Den første regelen tilsvarer restriksjonene (5.5) - (5.7), som var hensiktsmessige i tilfellet med symmetrisk log-lineær modell. I dette tilfelle har imidlertid regel nr. 2 visse regnetekniske fordeler, og vi kan da omskrive (5.16) til

$$(5.19) \quad \left\{ \begin{array}{l} p_{gj} = \frac{\exp(x_{-g}^j \beta_j)}{r-1 + \sum_{h=1}^r \exp(x_{-g}^h \beta_h)} \quad (j=1,2,\dots,r-1) \\ p_{gr} = \frac{1}{r-1 + \sum_{h=1}^r \exp(x_{-g}^h \beta_h)} \end{array} \right.$$

I spesialtilfellet $r=2$ (binær avhengig variabel) degenerer modellen som seg hør og bør til den binomiske logit-modellen som vi betraktet i avsnitt 5.1, jfr. likning (5.1):

$$(5.20) \quad p_{g1} = \frac{\exp(x_g' \beta_1)}{1 + \exp(x_g' \beta_1)} = [1 + \exp(-x_g' \beta_1)]^{-1}$$

Også denne modellen følger ved å betrakte de betingede sannsynligheter i den generelle log-lineære modell (5.4). Av (5.19) kan vi utlede

$$(5.21) \quad \log(p_{gh}/p_{gi}) = x_g'(\beta_h - \beta_i).$$

I den log-lineære modellen har vi, analogt med (5.11) i binær-tilfellet,

$$(5.22) \quad p_{gi} = P(a=i|(b,c)=g) = P(a=i|(b,c)=(j,k)) \\ = \frac{p_{ijk}}{\sum_h p_{hjk}}$$

slik at

$$(5.23) \quad \log(p_{gh}/p_{gi}) = \log p_{hjk} - \log p_{ijk} \\ = (\lambda_h^a - \lambda_i^a) + (\lambda_{hj}^{ab} - \lambda_{ij}^{ab}) + (\lambda_{hk}^{ac} - \lambda_{ik}^{ac}) + (\lambda_{hjk}^{abc} - \lambda_{ijk}^{abc}).$$

I begge modeller er altså log-odds en lineær funksjon av parametrene. Det er en én-entydig sammenheng mellom parametersettet $\beta_1, \dots, \beta_{r-1}$ og de λ -parametrene i den log-lineære modell som gjelder den avhengige variable. I det første tilfellet har vi normert ved hjelp av basis-grupper, å det andre tilfellet ved å kreve at alle parametre summert over én grupperingsvei skal være lik null.

Den multinomiske logit-modell kan altså utledes som de betingede sannsynligheter i en generell log-lineær sannsynlighetsmodell. I denne forstand er den log-lineære modell mer generell enn logit-modellen - den sier noe om fordelingen til alle variablene og behandler disse helt symmetrisk. Logit-modellen derimot, betrakter én variabel som den avhengige, og gir fordelingen til denne variabelen betinget m.h.p. de øvrige variable.

I en annen forstand er imidlertid logit-modellen mer generell enn den log-lineære modell. Den log-lineære modell er en teori om den sannsynlighetsstrukturen som ligger bak en fler-veis kontingenstabell. Alle variablene er således forutsatt gruppert - de er målt på nominalnivå. Ser vi derimot på den multinomiske logit-modellen (5.16), så er det ingenting i veien for at variabel-vektoren x_g kan inneholde variable på høyeste målenivå. Ved å spesifisere x_g som bestående av en passende rekke dummy-variable, har vi sett at modellen kan nyttes til analyse av betingede sannsynligheter i et gruppert materiale. Modellens anvendbarhet stanser imidlertid ikke her. x_g kan gjerne bestå av bare kontinuerlige variable (målt på intervall- eller forholdstallsnivå), eller av en blanding variable på alle målenivå. Estimeringsprosessen blir gjerne mer arbeidskrevende rent regneteknisk jo høyere målenivå det er på variablene, men i prinsippet ligger det her ingen begrensninger i modellen.

5.3 Estimering ved to-trinns generaliserte minste kvadraters metode (2SGLS)

Med utgangspunkt i den multinomiske logit-modellen (5.19), la oss betrakte transformasjonen

$$(5.24) \quad \log(p_{gj}/p_{gr}) = x_g' \beta_j \quad (g=1,2,\dots,s; j=1,2,\dots,r-1)$$

og la oss definere "restleddet"

$$(5.25) \quad w_{gj} = \log(y_{gj}/y_{gr}) - \log(p_{gj}/p_{gr}).$$

Vi kan da skrive

$$(5.26) \quad f_{gj} = \log(y_{gj}/y_{gr}) = x'_{gj}\beta_j + w_{gj}$$

eller

$$(5.27) \quad \underline{f} = Z\underline{\beta} + \underline{w}$$

på matriseform, analogt med (4.20). Likning (5.27) uttrykker et sett av observerbare variable, nemlig log-oddsene \underline{f} , som en lineær funksjon av β -parametrene pluss et restledd. Dersom restleddene \underline{w} tilfredsstiller de vanlige kravene til en regresjonsmodell, kan vi bruke en versjon av minste kvadraters metode til estimering av $\underline{\beta}$.

Restleddet w_{gj} kan omskrives som følger:

$$(5.28) \quad \begin{aligned} w_{gj} &= \log \left[(p_{gj} + u_{gj}) / (p_{gr} + u_{gr}) \right] - \log(p_{gj}/p_{gr}) \\ &= \log \left[\frac{p_{gj}(1+u_{gj}/p_{gj})}{p_{gr}(1+u_{gr}/p_{gr})} \right] - \log \left[\frac{p_{gj}}{p_{gr}} \right] \\ &= \log(1+u_{gj}/p_{gj}) - \log(1+u_{gr}/p_{gr}). \end{aligned}$$

Restleddene u_{gj} er de samme som vi kjenner fra tidligere, med fordelingsegenskapene (4.13) - (4.15).

Ved å ta Taylor-ekspansjonen av funksjonen (5.28) kan vi finne (de asymptotiske) fordelingsegenskapene til w_{gj} :

$$(5.29) \quad w_{gj} = (u_{gj}/p_{gj} - u_{gr}/p_{gr}) + v_{gj}$$

der

$$(5.30) \quad v_{gj} = \sum_{h=2}^{\infty} \frac{1}{h} (-1)^{h-1} \left[\left(\frac{u_{gj}}{p_{gj}} \right)^h - \left(\frac{u_{gr}}{p_{gr}} \right)^h \right].$$

Leddet v_{gj} inneholder ledd av orden 2 og høyere i Taylor-rekken. v_{gj} er av størrelsesorden $\frac{1}{n_g}$ i sannsynlighet, dvs. at vi kan se bort fra v_{gj} uten at dette vil påvirke de asymptotiske egenskaper til estimatorene for β . (Se f.eks. Zellner og Lee (1965), eller Bishop et al. (1975), kap. 14). De asymptotiske egenskaper gjelder strengt tatt bare når n_g går mot uendelig. Dersom antallet observasjoner n_g er rimelig stort, kan vi forutsette at de asymptotiske egenskaper gjelder tilnærmet, altså at

$$(5.31) \quad w_{gj} \approx u_{gj}/p_{gj} - u_{gr}/p_{gr}.$$

Da har vi, ved å bruke (4.12) - (4.16),

$$(5.32) \quad E(w_{gj}) \approx 0$$

$$\begin{aligned}
 (5.33) \quad \text{var}(w_{gj}) &\approx p_{gj}^{-2} \text{var}(u_{gj}) + p_{gr}^{-2} \text{var}(u_{gr}) - 2p_{gj}^{-1}p_{gr}^{-1} \text{cov}(u_{gj}, u_{gr}) \\
 &= \frac{1-p_{gj}}{n_g p_{gj}} + \frac{1-p_{gr}}{n_g p_{gr}} + \frac{2}{n_g} \\
 &= n_g^{-1} (p_{gj}^{-1} + p_{gr}^{-1}) \quad (j=1,2,\dots,r-1)
 \end{aligned}$$

$$\begin{aligned}
 (5.34) \quad \text{cov}(w_{gj}, w_{gk}) &= E [w_{gj} \cdot w_{gk}] - E(w_{gj}) \cdot E(w_{gk}) \\
 &\approx E \left[\left(\frac{u_{gj}}{p_{gj}} - \frac{u_{gr}}{p_{gr}} \right) \left(\frac{u_{gk}}{p_{gk}} - \frac{u_{gr}}{p_{gr}} \right) \right] \\
 &= p_{gj}^{-1} p_{gk}^{-1} \text{cov}(u_{gj}, u_{gk}) - p_{gk}^{-1} p_{gr}^{-1} \text{cov}(u_{gk}, u_{gr}) - p_{gj}^{-1} p_{gr}^{-1} \text{cov}(u_{gj}, u_{gr}) + p_{gr}^{-2} \text{var}(u_{gr}) \\
 &= (n_g p_{gr})^{-1} \quad (g=1,2,\dots,s; j \neq k)
 \end{aligned}$$

$$(5.35) \quad \text{cov}(w_{gj}, w_{hk}) \approx 0, \quad (g \neq h).$$

Disse formlene for restleddsmomentene i den transformerte modell (5.26) overrasker vel først og fremst ved sin enkelhet. Det slår oss dessuten at det er en bestemt, uhyre enkel sammenheng mellom momentene til w og de tilsvarende elementer i den inverterte kovariansmatrisen Σ^{-1} , som gjelder for de "opprinnelige" restledd u :

$$\begin{aligned}
 (5.36) \quad \Sigma_w \equiv \text{Cov}(w) &\approx \begin{bmatrix} N^{-2}_{\Sigma} 11 & N^{-2}_{\Sigma} 12 & \dots & N^{-2}_{\Sigma} 1, r-1 \\ N^{-2}_{\Sigma} 12 & N^{-2}_{\Sigma} 22 & \dots & N^{-2}_{\Sigma} 2, r-1 \\ \vdots & \vdots & \ddots & \vdots \\ N^{-2}_{\Sigma} 1, r-1 & N^{-2}_{\Sigma} 2, r-1 & \dots & N^{-2}_{\Sigma} r-1, r-1 \end{bmatrix} \\
 &= (I_{r-1} \otimes N^{-2})_{\Sigma}^{-1} = M^{-2}_{\Sigma}^{-1}
 \end{aligned}$$

der Σ^{-1} og Σ^{jj} og Σ^{jk} er som i (4.52) - (4.54). Den inverse av Σ_w er dermed

$$(5.37) \quad \Sigma_w^{-1} = \left[(I_{r-1} \otimes N^{-2})_{\Sigma}^{-1} \right]^{-1} = \Sigma (I_{r-1} \otimes N^2) = \Sigma M^2.$$

I spesialtilfellet $N=I$ (individdata) har vi endog

$$(5.38) \quad \Sigma_w^{-1} = \Sigma = \Sigma_0.$$

2SGLS-estimatoren til β i modellen (5.27) er således tilnærmet gitt ved

$$(5.39) \quad \underline{d} = (Z' \hat{\Sigma}_w^{-1} Z)^{-1} Z' \hat{\Sigma}_w^{-1} \underline{f}.$$

Her er $\hat{\Sigma}_w^{-1}$ den konsistente estimator av Σ_w^{-1} , dannet ved å erstatte p_{gj} med y_{gj} i uttrykkene for Σ_w^{-1} . \underline{d} har asymptotisk fordeling gitt ved

$$(5.40) \quad \sqrt{n} (\underline{d} - \beta) \xrightarrow{L} N(0, \lim_{n \rightarrow \infty} n (Z' \hat{\Sigma}_w^{-1} Z)^{-1}).$$

En konsistent estimator av kovariansmatrisen i denne fordelingen er

$$(5.41) \quad n\hat{\Sigma}_d = n(Z'\hat{\Sigma}_W^{-1}Z)^{-1} = n(Z'M^2\hat{\Sigma}Z)^{-1}.$$

Ifølge setning A.6 har vi da, under nullhypotesen $H_0 : K\beta = \kappa$ (4.46):

$$(5.42) \quad n(Kd-\kappa)'[K\hat{\Sigma}_dK']^{-1}(Kd-\kappa) \xrightarrow{L} \chi^2(\nu).$$

En asymptotisk test av H_0 på nivå α består dermed i å forkaste H_0 dersom observatoren (5.42) er større enn $(1-\alpha)$ -fraktilen i χ^2 -kvadratfordelingen med ν frihetsgrader.

2SGLS anvendt på den multinomiske logit-modell har i noen grad samme svakhet som i den lineære modell: for å kunne nytte y_{gj} som estimator av p_{gj} må vi bruke et gruppert materiale med rimelig store grupper, slik at ikke $y_{gj} = 0$ eller $y_{gj} = 1$.

For logit-modellen med grupperte data kan vi imidlertid anvende IGLS (jfr. avsnitt 4.3) uten å løpe oppi problemer med negative sannsynlighetsanslag. Framgangsmåten ved IGLS blir nøyaktig som beskrevet i avsnitt 4.3, bortsett fra at vi for hvert trinn t må regne oss tilbake til \hat{y}^t v.h.a. transformasjonen

$$(5.43) \quad \hat{y}_{gj}^t = \begin{cases} \frac{\exp(\hat{f}_{gj}^t)}{[1 + \sum_{k=1}^{r-1} \exp(\hat{f}_{gk}^t)]} & (j=1,2,\dots,r-1) \\ \frac{1}{[1 + \sum_{k=1}^{r-1} \exp(\hat{f}_{gk}^t)]} & (j=r) \end{cases}$$

(5.43) er den inverse transformasjon av (5.26) og gir estimatene av sannsynlighetene p_{gj} .

IGLS kan imidlertid ikke nyttes dersom vi har individdata, siden transformasjonen (5.26) da ikke gjelder.

Asymptotisk har IGLS- og 2SGLS-estimatorene samme fordeling.

5.4 Estimering ved sannsynlighetsmaksimering (ML)

5.4.1 Multinomisk logit-modell

"Log-likelihood"-funksjonen til den multinomiske logit-modell (5.19) fåes ved innsetting av (5.19) i (4.83). Dette gir:

$$(5.44) \quad \log L = \sum_{g=1}^s \left\{ \sum_{h=1}^{r-1} n_{gh} x_{gh}' \beta_h - n_g \cdot \log \left[1 + \sum_{h=1}^{r-1} \exp(x_{g-h}' \beta_h) \right] \right\} + \text{konstant}.$$

Ved derivasjon:

$$(5.45) \quad \begin{aligned} \frac{\partial \log L}{\partial \beta_{jk}} &= \sum_{g=1}^s \left\{ n_{gj} x_{gk} - n_g x_{gk} \cdot \frac{\exp(x_{g-j}' \beta_j)}{[1 + \sum_{h=1}^{r-1} \exp(x_{g-h}' \beta_h)]} \right\} \\ &= \sum_{g=1}^s \left\{ n_g x_{gk} (y_{gj} - p_{gj}) \right\}. \end{aligned}$$

ML-estimatorene finnes ved løsning av det ikke-lineære likningssystemet

$$(5.46) \quad \frac{\partial \log L}{\partial \beta_{jk}} = 0 \quad (j=1,2,\dots, r-1; k=1,2,\dots, m)$$

med hensyn på alle β_{jk} .

De annenderiverte av log L er gitt ved

$$(5.47) \quad \frac{\partial^2 \log L}{\partial \beta_{jk} \partial \beta_{ih}} = \sum_{g=1}^s \left[-n_g x_{gk} \frac{\partial p_{gj}}{\partial \beta_{ih}} \right] \\ = \sum_{g=1}^s \left[-n_g x_{gk} x_{gh} (\delta_{ij} p_{gi} - p_{gi} p_{gj}) \right],$$

der δ_{ij} betegner Kronecker-delta. Matrisen av annenderiverte består av $(r-1)^2$ blokker av dimensjon (mxm) . Element (k,h) i blokk (i,j) er gitt ved (5.47). På matriseform kan blokk (i,j) skrives som $-X'N^2_{\Sigma_{ij}}X$, der Σ_{ij} er gitt ved (4.27) og (4.28). Hele matrisen av annenderiverte kan dermed skrives

$$(5.48) \quad \frac{\partial^2 \log L}{\partial \beta \partial \beta'} = -(I_{r-1} \otimes X')(I_{r-1} \otimes N^2)\Sigma(I_{r-1} \otimes X) \\ = -Z'M^2\Sigma Z = -nZ'Q_n\Sigma_0 Z = -Z'\Sigma_W^{-1}Z.$$

Informasjonsmatrisen er følgelig

$$(5.49) \quad \Psi = -E \left[\frac{\partial^2 \log L}{\partial \beta \partial \beta'} \right] = Z'\Sigma_W^{-1}Z = nZ'Q_n\Sigma_0 Z.$$

En konsistent estimator av informasjonsmatrisen er gitt ved

$$(5.50) \quad \tilde{\Psi} = Z'\tilde{\Sigma}_W^{-1}Z = Z'M^2\tilde{\Sigma}Z$$

der $\tilde{\Sigma}_W^{-1}$ og $\tilde{\Sigma}$ betegner matrisene Σ_W^{-1} og Σ , innsatt for ML-estimatene \hat{p}_{gj} istedenfor p_{gj} .

La $\hat{\beta}$ betegne ML-estimatoren av β . Siden ML-estimatoren er asymptotisk effisient (jfr. avsnitt A.3.8-9) vet vi nå at

$$(5.51) \quad \sqrt{n}(\hat{\beta} - \beta) \xrightarrow{L} N(0, \lim_{n \rightarrow \infty} n\Psi^{-1}).$$

Hypoteser av formen (4.46) kan testes asymptotisk v.h.a. relasjonen

$$(5.52) \quad (K\hat{\beta} - \kappa)' [K\tilde{\Psi}^{-1}K']^{-1} (K\hat{\beta} - \kappa) \xrightarrow{L} \chi^2(\nu),$$

som gjelder under H_0 . Venstresiden i (5.52) har m.a.o. samme asymptotiske fordeling som kji-kvadrat-observatoren basert på sannsynlighetskvoten, idet (jfr. avsnitt A.3.9)

$$(A.61) \quad -2 \log(L_0/L_1) \xrightarrow{L} \chi^2(\nu).$$

L_0 og L_1 står her for "likeligheten" minimert under henholdsvis H_0 og H_1 .

Dersom H_1 er den minst restriktive av alle hypoteser, dvs. den mettede modell, blir denne observatoren

$$(5.53) \quad -2 \log(L_0/L_1) = -2 \sum_{g=1}^s \sum_{j=1}^r n_{gj} \log \hat{p}_{gj} + 2 \sum_{g=1}^s \sum_{j=1}^r n_{gj} \log(n_{gj}/n_g) \\ = 2 \sum_{g=1}^s \sum_{j=1}^r n_g y_{gj} \log(y_{gj}/\hat{p}_{gj}).$$

Her står \hat{p}_{gj} for ML-estimatene under H_0 . ML-estimatene i den mettede modell (H_1) blir selvsagt lik de observerte verdier $y_{gj} = n_{gj}/n_g$.

5.4.2 Symmetrisk log-lineær formulering

Betrakt igjen den symmetriske log-lineære formulering (5.4) - (5.7). Sambandet mellom de ubetingede sannsynlighetene p_{ijk} i den symmetriske 3-veis modell, og de tilsvarende betingede sannsynligheter p_{gi} i den assymmetriske formulering, er gitt ved (5.22). Dersom vi holder oss til de ubetingede sannsynlighetene p_{ijk} , og de tilsvarende observerte relative og absolutte frekvenser y_{ijk} og n_{ijk} , kan kji-kvadratobservatoren basert på sannsynlighetsknoten (5.53) omskrives til

$$(5.54) \quad G^2 = -2 \log(L_0/L_1) = 2 \sum_{i=1}^{s_a} \sum_{j=1}^{s_b} \sum_{k=1}^{s_c} n_{ijk} \log(y_{ijk}/\hat{p}_{ijk})$$

$$= 2 \sum_{i=1}^{s_a} \sum_{j=1}^{s_b} \sum_{k=1}^{s_c} n_{ijk} \log(n_{ijk}/\hat{n}_{ijk}).$$

$\hat{n}_{ijk} = n \hat{p}_{ijk}$ står her for sannsynlighetsmaksimeringsestimaterne av np_{ijk} , som altså finnes ved minimering av G^2 . Det er en én-entydig sammenheng mellom ML-estimaterne \hat{p}_{ijk} av p_{ijk} på den ene side og ML-estimaterne $\hat{\lambda}_i^a, \dots, \hat{\lambda}_{ijk}^{abc}$ av λ -parametrene (5.4) på den annen side.

Likning (5.54) gir, i likhet med (5.53), kji-kvadratobservatoren for testing mot den mettede modell. Mer generelt, dersom vi lar $\hat{n}_{ijk}^{(f)}$ betegne ML-estimatorene beregnet under H_f , og lar $G^2(H_0|H_1)$ betegne kji-kvadratobservatoren basert på sannsynlighetsknoten for testing av H_0 mot H_1 , så er

$$(5.55) \quad G^2(H_0|H_1) = -2 \log(L_0/L_1)$$

$$= 2 \sum_{i,j,k} n_{ijk} \log(\hat{n}_{ijk}^{(1)}/\hat{n}_{ijk}^{(0)})$$

$$= G^2(H_0) - G^2(H_1).$$

$G^2(H_f)$ står her for kji-kvadratobservatoren for testing av H_f mot den mettede modell ($f=0,1$), dvs.

$$(5.56) \quad G^2(H_f) = 2 \sum_{i,j,k} n_{ijk} \log(n_{ijk}/\hat{n}_{ijk}^{(f)}).$$

Når dette ikke gir grunnlag for misforståelser, vil vi for korthets skyld skrive bare G^2 istedenfor $G^2(H_f)$.

Celleestimaterne \hat{n}_{ijk} , og dermed observatoren G^2 , vil avhenge av hvilke bibetingelser vi legger på λ -ene. Dersom alle λ -er tillates å variere fritt (ingen bibetingelser), har vi som nevnt den mettede modell. I så fall blir $\hat{n}_{ijk} = n_{ijk}$ og $G^2 = 0$. Det kan vises at vi alltid vil få $G^2 \geq 0$, selv om bidraget fra enkelte celler (der hvor $n_{ijk} < \hat{n}_{ijk}$) vil være negativt.

5.5 Hierarkiske hypoteser for den log-lineære modell

Den log-lineære modellen har en del spesielle egenskaper som gjør den særlig attraktiv som analyseverktøy. Dersom vi setter opp hypotesen

$$(5.57) \quad H_1 : \lambda_{ijk}^{abc} = 0 \quad \text{for alle } i,j,k$$

er dette nemlig det samme som å si at

$$(5.58) \quad \frac{p_{ijk}p_{igk}}{p_{fjk}p_{fgk}} = \frac{p_{ijh}p_{igh}}{p_{fjh}p_{fgh}} \quad \text{for alle } f,g,h,i,j,k \quad (\text{se Haldorsen (1977)}).$$

Venstresiden i (5.58) gir uttrykk for samspillsstrukturen mellom variablene a og b i den to-veistabellen vi får ved bare å se på de celler der $c=k$. Det er altså den betingede to-faktor-sammenheng mellom a og b, gitt $c=k$. Høyresiden i (5.58) beskriver helt tilsvarende den betingede sammenheng mellom a og b, gitt $c=h$. Relasjon (5.58) sier altså at dette betingede samvariasjonsforholdet er uavhengig av c. Vi kan si at (5.57) er en hypotese om uavhengighet på 3-faktornivået.

Likning (5.58) behandler tilsynelatende variabel c assymmetrisk i forhold til de andre to grupperingsveier: tredje indeks er hele veien lik k på venstresiden og h på høyresiden. I virkeligheten er imidlertid (5.58) en helt symmetrisk hypotese om tre-faktorforholdet (a,b,c). Ved å ommøblere likningen på passende måte vil vi se at den også innebærer at samspillet (b,c) er uavhengig av a, og at samspillet (a,c) er uavhengig av b.

La oss nå betrakte hypotesen

$$(5.59) \quad H_2 : \lambda_{ijk}^{abc} = \lambda_{ij}^{ab} = 0 \text{ for alle } i,j,k.$$

I tillegg til H_1 forutsetter vi altså nå at to-faktoreffektene mellom a og b er lik 0. Denne hypotesen innebærer, i tillegg til (5.58), at

$$(5.60) \quad p_{ijk} = \frac{p_{i+k} p_{+jk}}{p_{++k}} \text{ for alle } i,j,k,$$

der $p_{i+k} = \sum_j p_{ijk}$, $p_{+jk} = \sum_i p_{ijk}$ og $p_{++k} = \sum_{ij} p_{ijk}$. Hver indeks som erstattes med plusstegn betyr at vi betrakter marginalen oppnådd ved å summere over vedkommende indeks.

Likning (5.60) har en naturlig tolkning som betinget uavhengighet mellom a og b, gitt $c=k$. Dette innses vi lett ved å omskrive (5.60) til

$$(5.61) \quad \frac{p_{ijk}}{p_{++k}} = \frac{p_{i+k}}{p_{++k}} \cdot \frac{p_{+jk}}{p_{++k}} \text{ for alle } i,j,k,$$

dvs.

$$(5.62) \quad P(a=i \text{ og } b=j | c=k) = P(a=i | c=k) \cdot P(b=j | c=k) \text{ for alle } i,j,k.$$

Denne betingede uavhengighet gjelder vel å merke for alle mulige verdier (k) av c.

H_1 sier at to-faktorsammenhengene (a,b), (a,c) og (b,c) ikke endrer seg med verdien på den tredje variabelen. H_2 sier i tillegg at for paret (a,b) er denne samvariasjonen lik null for alle verdier av c.

La oss likevel for ordens skyld understreke at (5.60) ikke er det samme som marginal uavhengighet mellom a og b, som vi eventuelt ville skrive slik:

$$(5.63) \quad p_{ij+} = p_{i++} p_{+j+}$$

eller

$$(5.64) \quad P(a=i \text{ og } b=j) = P(a=i) \cdot P(b=j).$$

Betrakt nå hypotesen

$$(5.65) \quad H_3 : \lambda_{ijk}^{abc} = \lambda_{ij}^{ab} = \lambda_{ik}^{ac} = 0 \text{ for alle } i,j,k.$$

I tillegg til H_2 forutsetter vi her betinget uavhengighet mellom a og c, gitt b. Det vil si at det nå bare er b og c som er innbyrdes avhengige. Både b og c er hver for seg uavhengige av a:

$$(5.66) \quad p_{ijk} = p_{i++}p_{+jk} \text{ for alle } i,j,k.$$

Hypotesen av total marginal uavhengighet mellom alle tre grupperingsveier kan skrives

$$(5.67) \quad H_4 : \lambda_{ijk}^{abc} = \lambda_{ij}^{ab} = \lambda_{ik}^{ac} = \lambda_{jk}^{bc} = 0 \text{ for alle } i,j,k$$

eller

$$(5.67) \quad p_{ijk} = p_{i++}p_{+j+}p_{++k} \text{ for alle } i,j,k.$$

Her er alle to-faktor- og høyere ordens effekter satt lik null.

Den aller mest restriktive hypotesen er den der også alle en-faktoreffektene λ_i^a , λ_j^b og λ_k^c settes lik null:

$$(5.69) \quad H_5 : \lambda_{ijk}^{abc} = \lambda_{ij}^{ab} = \lambda_{ik}^{ac} = \lambda_{jk}^{bc} = \lambda_i^a = \lambda_j^b = \lambda_k^c = 0 \text{ for alle } i,j,k.$$

H_5 innebærer at

$$(5.70) \quad p_{ijk} = \frac{1}{s_a s_b s_c} \text{ for alle } i,j,k,$$

m.a.o. at alle celler i tre-veis-tabellen har like stor sannsynlighet.

Vi kan også tenke oss hypoteser av typen

$$(5.71) \quad H_6 : \lambda_{ijk}^{abc} = \lambda_{ij}^{ab} = \lambda_{ik}^{ac} = \lambda_i^a = 0 \text{ for alle } i,j,k$$

$$(5.72) \quad p_{ijk} = \frac{1}{s_a} p_{+jk} \text{ for alle } i,j,k,$$

$$(5.73) \quad H_7 : \lambda_{ijk}^{abc} = \lambda_{ij}^{ab} = \lambda_{ik}^{ac} = \lambda_{jk}^{bc} = \lambda_i^a = 0 \text{ for alle } i,j,k$$

$$(5.74) \quad p_{ijk} = \frac{1}{s_a} p_{+j+}p_{++k} \text{ for alle } i,j,k,$$

eller

$$(5.75) \quad H_8 : \lambda_{ijk}^{abc} = \lambda_{ij}^{ab} = \lambda_{ik}^{ac} = \lambda_{jk}^{bc} = \lambda_i^a = \lambda_j^b = 0 \text{ for alle } i,j,k$$

$$(5.76) \quad p_{ijk} = \frac{1}{s_a s_b} p_{++k} \text{ for alle } i,j,k.$$

Hver av disse hypotesene, om at bestemte λ -parametre er lik null, kan altså tolkes som en eller annen type (betinget eller marginal) uavhengighet mellom variablene. Tilsvarende egenskaper kan den lineære sannsynlighetsmodellen ikke framvise.

Alle de nevnte hypotesene er eksempler på såkalt hierarkiske hypoteser. Med dette mener vi at dersom t-faktoreffektene mellom en bestemt gruppe variabler er lik null, må også alle (t+1)- og høyere ordens effekter som omfatter disse samme variablene, være satt lik null. F.eks., dersom $\lambda_i^a = 0$, så må også $\lambda_{ij}^{ab} = 0$, $\lambda_{ik}^{ac} = 0$ og $\lambda_{ijk}^{abc} = 0$ (jfr.(5.71)).

Vi skal se litt nærmere på hvordan kji-kvadratobservatorene G^2 kan nyttes til testing av slike hierarkiske hypoteser. La H_f og H_g være to forskjellige hierarkiske hypoteser slik at mengden av likninger som trenges for å spesifisere H_f er en ekte undermengde av likningssettet for H_g . Vi sier at H_f er "inneholdt i H_g " ($H_f \subset H_g$). Alternativt kan vi si at parameterrommet under H_g (Ω_g) er inneholdt i parameterrommet under H_f , dvs. $\Omega_g \subset \Omega_f$. Alt det som blir påstått av H_f blir m.a.o. også påstått av H_g . H_g er en "ekte strengere hypotese" enn H_f - den påstår det samme pluss litt til. Likningene (5.57) - (5.76) gir tallrike eksempler på slike par av hypoteser. H_1 er f.eks. inneholdt i alle de andre hypotesene. H_2 er inneholdt i H_3 , som igjen er inneholdt i H_4 , osv.:

$$(5.77) \quad H_1 \subset H_2 \subset H_3 \subset H_4 \subset H_7 \subset H_8 \subset H_5.$$

Merk at

$$(5.78) \quad H_f \subset H_g \Rightarrow G^2(H_f) \leq G^2(H_g),$$

slik at (5.77) medfører

$$(5.79) \quad 0 \leq G^2(H_1) \leq G^2(H_2) \leq G^2(H_3) \leq G^2(H_4) \leq G^2(H_7) \leq G^2(H_8) \leq G^2(H_5).$$

Under H_f er $G^2(H_f)$ asymptotisk kji-kvadratfordelt med v_f frihetsgrader. v_f er bestemt som antall uavhengige celler i tabellen minus antall uavhengige parametre. Alternativt kan vi regne ut v_f som antall uavhengige likninger vi trenger for å spesifisere H_f med utgangspunkt i den mettede modell.

Dersom vi vet at H_f er sann, og ønsker å teste H_g ($H_f \subset H_g$), kan vi bruke observatoren

$$(5.80) \quad G^2(H_g|H_f) = G^2(H_g) - G^2(H_f),$$

jfr. (5.55). Det kan nemlig vises at, under H_g , så gjelder nå

$$(5.81) \quad G^2(H_g|H_f) \xrightarrow{L} \chi^2(v_g - v_f).$$

Observatoren (5.80) er m.a.o. under H_g asymptotisk kji-kvadratfordelt med $v_g - v_f$ frihetsgrader. Vi betinger m.h.p. H_f , og oppnår en sterkere test ved slik å utnytte informasjonen om at H_f er sann.

I avsnitt 4.7 (setning 4.2) utledet vi en helt analog testmetode for tilfellet med lineær sannsynlighetsmodell, da basert på Neyman kji-kvadrat-observatoren (χ_N^2). I forhold til den lineære modellen har imidlertid den log-lineære modell det fortrinn at når alle variable er kvalitative, kan enhver hierarkisk hypotese tolkes som uavhengighet mellom bestemte variable på et eller annet nivå.

I avsnitt 9.4 vil vi komme tilbake til hvordan dette testapparatet kan brukes til å analysere et empirisk materiale. Forøvrig vil vi her ikke gå nærmere inn på teorien for log-lineære modeller. Det eksisterer en omfattende litteratur på området. Særlig lettfattelige framstillinger er gjort av Haldorsen (1976 og 1977). Bishop et al. (1975) og Goodman (1970 og 1971) er mindre lettleste, men til gjengjeld svært grundige "kokebøker". Goodman er mannen bak programmet ECTA, som er beskrevet på norsk av Clausen (1977).

5.6 Estimering ved minimum logit kji-kvadrat ($\min \chi_L^2$)

For tilfellet med dikotom avhengig variabel ($r=2$) og grupperte observasjoner foreslo Berkson (1944, 1953 og 1955) en estimeringsmetode han kalte minimum logit-kji-kvadrat. Estimatorene oppnås ved minimering av uttrykket

$$(5.82) \quad \begin{aligned} \chi_L^2 &= \sum_{g=1}^s n_g y_{g1} (1-y_{g1}) \left[\log\left(\frac{y_{g1}}{1-y_{g1}}\right) - \log\left(\frac{p_{g1}}{1-p_{g1}}\right) \right]^2 \\ &= \sum_{g=1}^s n_g y_{g1} (1-y_{g1}) \left[f_{g1} - x'_{g1} \beta_1 \right]^2 \end{aligned}$$

med hensyn på β . Størrelsen χ_L^2 kalles logit kji-kvadrat-observatoren og er asymptotisk kji-kvadrat-fordelt med $\nu = s-m$ frihetsgrader, dersom (5.19) er den sanne modell. Her er som før s antall celler i kryssgrupperingen av alle uavhengige variable, og m antall elementer i parametervektoren β . Uttrykket (5.82) er det log-lineære motstykket til Neyman kji-kvadrat-observatoren (4.92), som gjaldt for den lineære sannsynlighetsmodell.

Også min χ_L^2 -metoden lar seg generalisere til det multinomiske tilfellet, men disse uttrykkene blir nokså kompliserte. χ_L^2 -observatoren har derfor stort sett bare blitt anvendt på det binomiske tilfellet.

På samme måte som min χ_N^2 er ekvivalent med 2SGLS i den lineære modell (jfr. avsnitt 4.7), så er min χ_L^2 ekvivalent med 2SGLS i den binomiske logit-modell. I spesialtilfellet $r=2$ reduseres nemlig (5.33) til

$$(5.83) \quad \text{var } w_{g1} = n_g^{-1} [p_{g1}^{-1} + (1-p_{g1})^{-1}] = [n_g p_{g1} (1-p_{g1})]^{-1}.$$

Minimering av (5.82) innebærer således at hver observasjon f_{g1} veies med et konsistent estimat av den inverse av standardavviket. Dette er nettopp hva 2SGLS går ut på i det dikotome tilfellet.

Både 2SGLS (=min χ_L^2) og ML er asymptotisk effisiente (Neyman (1949)). Når antallet observasjoner blir svært stort gir altså begge metoder like gode estimater.

5.7 Sammenlikning mellom 2SGLS og ML

Teorien forteller oss at ML-estimatorene er asymptotisk optimale; de er BAN ("best asymptotically normal"), dvs. at den asymptotiske kovariansmatrisen til ML-estimatoren faller sammen med den inverse av informasjonsmatrisen (jfr. (5.51)).

Sammenlikning av (5.40) og (5.49) forteller oss imidlertid at 2SGLS- (og IGLS-) estimatorene har samme egenskap: de konvergerer til samme multinormale fordeling som ML-estimatoren. Når utvalget vokser mot uendelig, spiller det liten rolle hvilken metode vi velger.

Hvilken metode er best i små utvalg? Begge estimatorene er generelt forventingsskjeve. Bortsett fra dette er de teoretiske fordelingsegenskapene ikke kjent for noen av metodene. Det foreligger imidlertid en del Monte-Carlo-studier o.l. som kaster lys over dette spørsmålet.

Først ute var Berkson (1955 a) for tilfellet $r=2$ og $m=2$ - binomisk logit-modell med konstantledd og én variabel. Han undersøkte de eksakte fordelingsegenskapene til ML- og 2SGLS- (= min χ_L^2 -) estimatorene for et tenkt utvalg på 30 observasjoner, 10 fra hver av 3 mulige verdier på variabelen x_1 . Berkson fant her det overraskende resultat at 2SGLS var best etter praktisk talt alle kriterier: skjevhet, varians eller bruttovarians, bortsett fra at konstantleddet β_1 for de fleste sanne verdier av β_1 ble estimert med mindre skjevhet av ML.

Berkson fant også et annet meget interessant resultat: I tilfellet med bare én ukjent parameter (konstantleddet β_1) var Cramér-Raos grense for bruttovariansen betydelig lavere for 2SGLS enn for ML. Ikke nok med det, 2SGLS-estimatoren ble funnet å ha lavere bruttovarians enn noen forventningsrett observator kan ha, ifølge Cramér-Rao's lov (A.51). (Om denne, se avsnitt A.3.7 i Appendiks.)

Dette paradokset har sin forklaring i følgende forhold. Betrakt en estimator $a(x)$ av en parameter θ . La $\alpha(\theta)$ være forventningen av $a(x)$ og la $\alpha'(\theta)$ være den deriverte av α m.h.p. θ . Cramér-Raos lov for bruttovariansen (bvar) kan da skrives

$$(A.54) \quad \text{bvar}_\theta(a(x)) \geq \frac{[\alpha'(\theta)]^2}{\psi(\theta)} + s_\theta^2,$$

der $\psi(\theta)$ er informasjonstallet og s_θ er skjevheten av estimatoren $a(x)$ m.h.p. parameteren θ . Den deriverte $\alpha'(\theta)$ kan skrives

$$(5.84) \quad \alpha'(\theta) = 1 + \frac{\partial s_\theta}{\partial \theta}.$$

Dersom $\partial s_{\theta} / \partial \theta$ er negativ, kan det tenkes at

$$(5.85) \quad \left[1 + \frac{\partial s_{\theta}}{\partial \theta} \right]^2 / \psi(\theta) + s_{\theta}^2 < 1 / \psi(\theta),$$

altså at estimatoren med forventning $\alpha(\theta) \neq \theta$ kan ha lavere grense for bruttovariansen enn den inverse av informasjonstallet, som gjelder for en forventningsrett estimator.

Berkson finner at nettopp dette fenomenet gjør seg gjeldende for 2SGLS-estimatoren. Skjevheten til 2SGLS-estimatoren har motsatt fortegn av den sanne parameteren β_1 , og den deriverte $\partial s_{\beta_1} / \partial \beta_1$ er overalt negativ. For ML-estimatoren, derimot, er $\partial s_{\beta_1} / \partial \beta_1$ positiv. Cramér-Rao's grense for bruttovariansen til 2SGLS-estimatoren er derfor mindre enn $1/\psi(\theta)$, mens den for ML-estimatoren er større enn $1/\psi(\theta)$. Til tross for at ML-estimatoren kommer nærmere sin nedre grense enn 2SGLS-estimatoren gjør, blir bruttovariansen lavest for 2SGLS-estimatoren.

Som rosinen i pølsa viser Berkson at det er mulig å ytterligere forbedre 2SGLS-estimatoren gjennom en prosess han kaller for "Rao-Blackwellisering". Det såkalte Rao-Blackwell-teoremet sier følgende. La $a(\underline{x})$ være en estimator for θ med endelig varians, og la $t(\underline{x})$ være en suffisient observator for θ . Da er

$$(5.86) \quad a^*(\underline{x}) = E[a(\underline{x}) | t(\underline{x})]$$

en minst like god estimator som $a(\underline{x})$, i den forstand at $E[a(\underline{x})] = E[a^*(\underline{x})]$ og

$$(5.87) \quad \text{bvar}_{\theta} a^*(\underline{x}) \leq \text{bvar}_{\theta} a(\underline{x}).$$

(5.87) gjelder som ekte ulikhet hvis og bare hvis $a^*(\underline{x}) \neq a(\underline{x})$. (En suffisient observator $t(\underline{x})$ er en observator som inneholder all den informasjon om en parameter (θ) som det er mulig å trekke ut av utvalgsobservasjonene (\underline{x}), i den forstand at den betingede fordeling av \underline{x} , gitt $t(\underline{x})$, ikke avhenger av θ . For en nærmere redegjørelse for både suffisiens-begrepet og Rao-Blackwell-teoremet viser vi til Bickel og Doksum (1977).)

Ved å ta den betingede forventning av $a(\underline{x})$ m.h.p. den suffisiente observator får vi altså en ny estimator som er minst like god som den gamle. Dersom $a(\underline{x})$ allerede i utgangspunktet er en funksjon av den suffisiente observator, blir nødvendigvis $a^*(\underline{x}) = a(\underline{x})$, og vi får ingen forbedring. Estimatorer som ikke kan uttrykkes som en funksjon av den suffisiente observator, kan vi imidlertid i prinsippet forbedre ved å anvende Rao-Blackwell-teoremet.

Mens ML-estimatoren er en funksjon av den suffisiente observator, er dette ikke tilfelle for 2SGLS-estimatoren i logit-modellen. "Rao-Blackwellisering" gir derfor forbedring av 2SGLS-estimatoren, men ikke av ML-estimatoren. Berkson beregner den "Rao-Blackwelliserte" 2SGLS-estimator og finner at denne faktisk når Cramér-Rao's nedre grense for bruttovariansen.

Her må vi likevel slå litt kaldt vann i blodet og presisere at "Rao-Blackwellisering" ikke er gratis. Utrekning av den betingede forventning (5.86) er generelt meget arbeidskrevende, hvis det i det hele tatt er mulig å finne en metode. Den "Rao-Blackwelliserte" estimator savner derfor en av 2SGLS-estimatorens viktigste egenskaper: enkelhet i utregningen.

Berksons undersøkelse av tilfellet med bare én eller to ukjente parametre og to mulige utfall er selvsagt for lite til at vi kan trekke helt generelle konklusjoner om ML- og 2SGLS-estimatorenes relative effisiens. Nyere studier (se f.eks. Berkson (1980) og Amemiya (1980)) bekrefter likevel Berksons tidligere resultater. Alle disse studiene er i alle fall egnet til å illustrere to viktige poeng: for det første at den asymptotisk effisiente observator stundom oppfører seg heller dårlig i endelige utvalg, og for det andre at vi ikke bør begrense oppmerksomheten til klassen av forventningsrette observatorer.

Nyere studier er gjort av McFadden (1974) og av Domencich og McFadden (1975). McFadden (1974) undersøker egenskapene til ML-estimatoren v.h.a. Monte-Carlo-eksperiment med fra 10 til 200 observasjoner. Skjevheten er stor (opptil 60 prosent) i de minste utvalgene, men reduseres nokså raskt når utvalget vokser: ved 200 observasjoner er skjevheten mellom 1 og 2 prosent, avhengig av den sanne parameterverdien.

Domencich og McFadden (1975) gjør en Monte-Carlo-studie for sammenlikning av ML- og 2SGLS-estimatorene i en situasjon der den uavhengige variable er kontinuerlig. For å beregne 2SGLS-estimatoren må den uavhengige variable reduseres til nominalnivå (grupperes), dette forsterker (den negative) skjevheten til 2SGLS-estimatoren.

Domencich og McFadden skiller mellom to tilfeller: (A) få sannsynligheter nær 0 eller 1, og (B) mange sannsynligheter nær 0 eller 1. I tilfelle (A) finner de at 2SGLS gir jevnt over noe lavere bruttovarians enn ML, men betydelig større skjevhet. I tilfelle (B) gir 2SGLS atskillig høyere bruttovarians enn ML unntatt for de minste utvalgene, og skjevheten til 2SGLS er ekstra stor.

Resultatene kan oppsummeres som følger: Dersom våre uavhengige variable i utgangspunktet er målt på nominal- eller ordinalnivå, er 2SGLS å foretrekke framfor ML. 2SGLS har antakelig bedre fordelingsegenskaper i mindre utvalg, og er likeverdig i store utvalg. Dersom de uavhengige variable foreligger målt for forholdstallsnivå, innebærer gruppering av dataene estimering ved 2SGLS et tap av effisiens. Fortsatt er imidlertid 2SGLS ikke vesentlig dårligere enn ML, unntatt i tilfellet med mange sannsynligheter nær 0 eller 1. Begge metoder gir imidlertid akseptable resultater forutsatt at utvalget ikke er av de aller minste. Valget mellom de to metoder blir derfor i stor grad et spørsmål om hensiktsmessighet og kostnad i den enkelte situasjon.

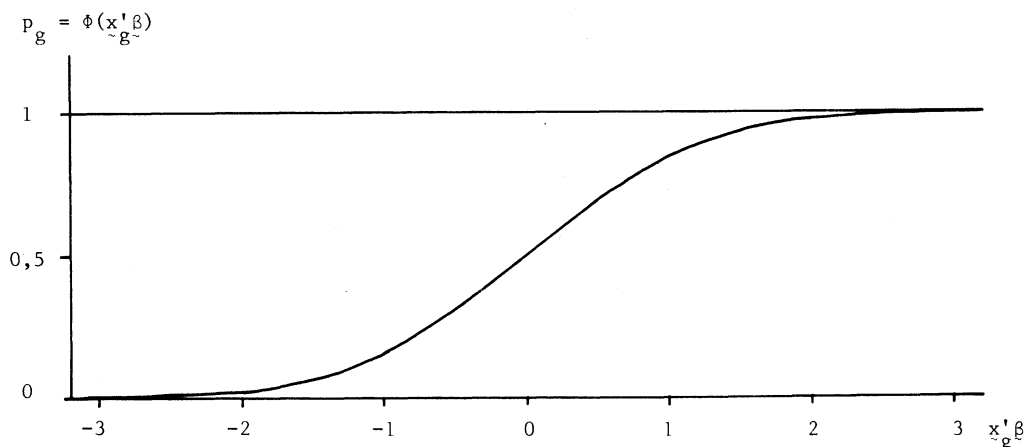
6. PROBIT-MODELLEN

Vi vil her kort omtale en konkurrent til den binomiske logit-modellen, nemlig den såkalte probit-modell, definert ved

$$(6.1) \quad P_i = \int_{-\infty}^{x_i^* \beta} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt = \Phi(x_i^* \beta).$$

Som funksjonsform $h(\cdot)$ velger en altså det normale integral, her kalt $\Phi(\cdot)$. Probit-funksjonen er framstilt grafisk i fig. 6.1.

Figur 6.1. Probit-funksjonen *The probit function*



Forløpet av probit-funksjonen er svært likt logit-funksjonen. Alle verdier av funksjonen ligger i området mellom 0 og 1. Probit-funksjonen er noe brattere enn logit-funksjonen rundt verdien $p = 0.5$, til gjengjeld er halene tynnere (kurven går raskere mot asymptotene $p_i=0$ og $p_i=1$).

La oss, som før, bruke notasjonen $p_i = P(y_i^*=1)$, der y_i^* er den avhengige binærvariable for et individ i . "Log-likelihood"-funksjonen til probit-modellen er gitt ved

$$(6.2) \quad \log L = \sum_{\{y_i^*=1\}} \log \phi(x_i^* \beta) + \sum_{\{y_i^*=0\}} \log [1 - \phi(x_i^* \beta)].$$

ML-estimatorene finnes (i prinsippet) ved løsning av det ikke-lineære likningssystemet

$$(6.3) \quad \frac{\partial \log L}{\partial \beta_j} = \sum_{\{y_i^*=1\}} \frac{\phi(x_i^* \beta) x_{ij}}{\phi(x_i^* \beta)} - \sum_{\{y_i^*=0\}} \frac{\phi(x_i^* \beta) x_{ij}}{1 - \phi(x_i^* \beta)} = 0 \quad (j=1, 2, \dots, m),$$

der $\phi(\cdot)$ er den deriverte av Φ , dvs. den normale tettheten:

$$(6.4) \quad \phi(x_i^* \beta) = (2\pi)^{-\frac{1}{2}} \exp[-\frac{1}{2}(x_i^* \beta)^2].$$

Både Φ og ϕ er relativt kompliserte funksjoner av $x_i^* \beta$, som dessuten inngår både i teller og nevner. ML-estimering av probit-modellen er derfor betydelig mer komplisert rent regneteknisk enn ML/logit. Systemet konvergerer ofte meget sent mot ML-estimatoren.

Monte-Carlostudier har vist at ML-estimatoren for probit-modellen kan ha temmelig stor forventningsskjevhet i mindre utvalg, og at skjevheten reduseres bare langsomt når en øker utvalgsstørrelsen (Goldfeld og Quandt (1972), s. 133).

For tilfellet med grupperte data har Berkson (1955 b og 1957) utviklet en estimeringsmetode som han kaller "minimum normit kji-kvadrat" ($\min \chi_{NO}^2$). Metoden er helt analog med minimum logit kji-kvadrat anvendt på logit-modellen (avsnitt 5.6):

La Φ^{-1} være den inverse funksjon av Φ , dvs. $\Phi^{-1}(p_g)$ er p_g -fraktilen i standard-normalfordelingen, og kalles "normit" av p_g . Definer restleddet

$$(6.5) \quad w_g = \Phi^{-1}(y_g) - \Phi^{-1}(p_g) = \Phi^{-1}(y_g) - x_g' \beta.$$

Analogt med utledningen i avsnitt 5.3 kan vi vise at $\text{var}(w_g)$ er tilnærmet gitt ved

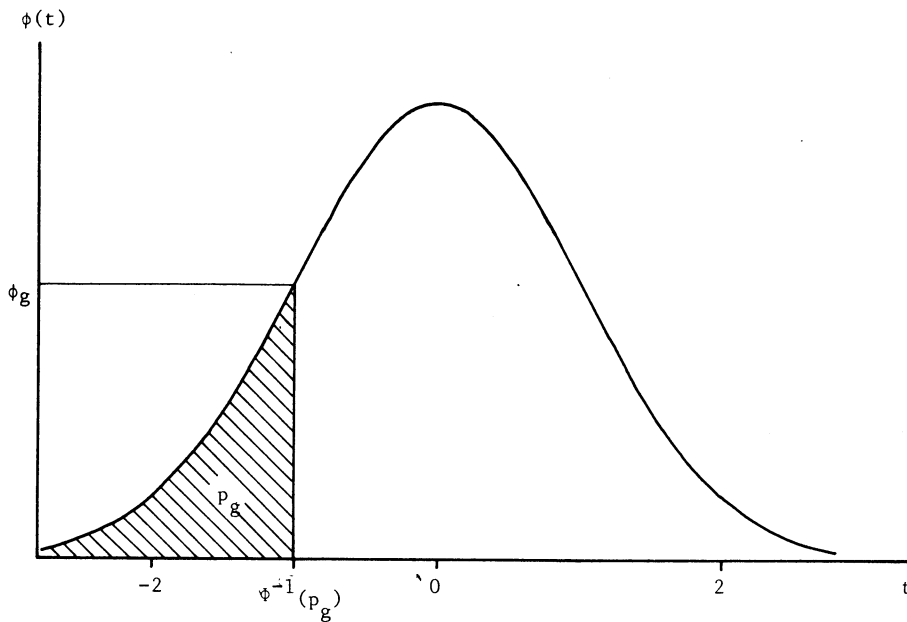
$$(6.6) \quad \text{var}(w_g) \approx \frac{p_g(1-p_g)}{n_g \phi_g^2}$$

der

$$(6.7) \quad \phi_g = \phi[\Phi^{-1}(p_g)] = (2\pi)^{-\frac{1}{2}} \exp[-\frac{1}{2}(x_g' \beta)^2].$$

ϕ_g er altså høyden på normalfordelingskurven i det punkt der den kumulative sannsynligheten er p_g , se fig. 6.2. Det skraverte arealet er $p_g = \Phi(x_g' \beta)$.

Figur 6.2. Sammenhengen mellom p_g og ϕ_g *The relationship between p_g and ϕ_g*



Et konsistent estimat av $\text{var}(w_g)$ kan vi få ved å sette inn y_g for p_g :

$$(6.8) \quad \widehat{\text{var}(w_g)} = \frac{y_g(1-y_g)}{n_g \hat{\phi}_g^2}$$

der

$$(6.9) \quad \hat{\phi}_g = \phi_g |_{p_g=y_g} = \phi[\phi^{-1}(y_g)].$$

2SGLS-estimering av probitmodellen er derfor algebraisk ekvivalent med minimering av "normit kji-kvadrat-observatoren"

$$(6.10) \quad x_{NO}^2 = \sum_{g=1}^s \frac{n_g \hat{\phi}_g^2}{y_g(1-y_g)} [\phi^{-1}(y_g) - x_{g\beta}^1]^2$$

m.h.p. parametrene β . x_{NO}^2 , innsatt for 2SGLS-estimaterne av β , er asymptotisk kji-kvadrat fordelt med det passende antall frihetsgrader, gitt at probit er den sanne modell.

Det største praktiske problem forbundet med denne metoden er i alminnelighet å regne ut $\phi^{-1}(y_g)$. Dersom denne funksjonen ikke finnes tilgjengelig i regneprogrambiblioteket har en imidlertid en rekke enkle approksimasjoner å ty til, se Abramowitz og Stegun (1968), Page (1977) eller Schmeiser (1979).

Berkson (1955 b og 1957) undersøkte og sammenliknet egenskapene til ML- og $\min x_{NO}^2$ (=2SGLS-) estimatene i tilfellet $m = 2$, der x_g består av én konstant og én variabel. Resultatene var svært parallele til det som ble funnet for logit-modellen, jfr. avsnitt 5.7. I mindre utvalg med grupperte data hadde $\min x_{NO}^2$ -estimatoren jevnt over mindre skjevhet enn ML-estimatoren, og vurdert etter brutto-variansen var $\min x_{NO}^2$ utvetydig best av de to. Asymptotisk er imidlertid begge metoder effisiente.

Vi vil generelt ikke anbefale bruk av probit-modellen, dersom det ikke er sterke a priori grunner til det. For praktiske formål er logit- og probit-funksjonene knapt til å skille fra hverandre (se fig. 10.1). Logit-funksjonen leder imidlertid til langt enklere regnearbeid. Estimering av probit-modellen ved sannsynlighetsmaksimering frarådes særlig sterkt, da fordelingsegenskapene i små utvalg er heller dårlige og løsningen av likningssystemet er svært ressurskrevende.

7. FØYNINGSMÅL FOR LINEÆRE OG LOG-LINEÆRE SANNSYNLIGHETSMODELLER

7.1 Determinasjonskoeffisienten (R^2) for den lineære modell

7.1.1 Generelt om OLS-estimering

Det vanlige mål for føyning i en regresjonsmodell er den såkalte determinasjonskoeffisienten, R^2 , definert ved

$$(7.1) \quad R^2 = 1 - \frac{\hat{u}'\hat{u}}{y'Ay}$$

der

$$(7.2) \quad \hat{u} = y - \hat{y}$$

er vektoren av residualer (av dimensjon $(t \times 1)$), og

$$(7.3) \quad A = (I_t - \frac{1}{t} i_t i_t')$$

A er en operator som omdanner elementene i en vektor til de tilsvarende avvik fra gjennomsnittet. Matrisen A har egenskapen

$$(7.4) \quad A^2 = A.$$

Nevneren $y'Ay$ er således den velkjente summen av de kvadrerte avvik fra gjennomsnittet for den avhengige variable, den "totale varians" i datasettet. Telleren $\hat{u}'\hat{u}$ er den "ikke forklarte varians".

Anvendt på vårt estimeringsproblem (se avsnitt 4.5) må determinasjonskoeffisienten skrives

$$(7.5) \quad R^2 = 1 - \frac{\hat{u}'M\hat{u}}{y'M^{\frac{1}{2}}AM^{\frac{1}{2}}y},$$

idet vi i realiteten har anvendt vanlig minste kvadraters metode på den transformerte modell (4.75). Ved hjelp av matrisen $M^{\frac{1}{2}}$ veies alle observasjonene med kvadratrotten av utvalgsstørrelsen n_g . Antall observasjoner er $s(r-1)$ og matrisen A er dermed

$$(7.6) \quad A = (I - \frac{1}{s(r-1)} j j')$$

Allerede ved å betrakte formelen (7.5) kan vi konstatere at dersom $r > 2$, er R^2 neppe et rimelig mål for føyning i denne type modell. Nevneren i (7.5) beregner nemlig avvik i forhold til (den veiede) gjennomsnittsandelen over alle $r-1$ mulige utfall. Et rimeligere føyningsmål ville beregne avvikene for hvert utfall separat.

I det følgende skal vi derfor i første rekke se på spesialtilfellet $r = 2$, altså med binær avhengig variabel.

Et annet poeng en kan merke seg ved formelen (7.5) er at de veiede residualene $M^{\frac{1}{2}}\hat{u}$ ikke nødvendigvis summerer seg til null, slik de vil gjøre ved OLS-estimering av en modell med konstantledd. Modellen (4.75) inneholder imidlertid teknisk sett ikke noe konstantledd; dette gjelder også for tilfellet $r = 2$.

En kunne derfor overveie å beregne avvik fra gjennomsnittet også for residualene \hat{u} , ved å sette $\hat{u}'M^{\frac{1}{2}}AM^{\frac{1}{2}}\hat{u}$ istedenfor $\hat{u}'M\hat{u}$. Theil (1971), s. 178 fraråder imidlertid dette, idet, som han sier:

"this would eliminate a nonzero average of the residuals, and such a nonzero average should be regarded as part of the poorness of fit".

Formelen (7.5) kan omformes som følger:

$$(7.7) \quad \hat{u}'M^{\frac{1}{2}}AM^{\frac{1}{2}}\hat{u} = \hat{u}'M\hat{u} - \frac{1}{s(r-1)} \hat{u}'M^{\frac{1}{2}}i i'M^{\frac{1}{2}}\hat{u}$$

$$(7.8) \quad = \hat{u}'M\hat{u} - \frac{1}{s(r-1)} (i'M^{\frac{1}{2}}\hat{u})^2$$

$$(7.9) \quad = \sum_{g=1}^s \sum_{j=1}^{r-1} n_g y_{gj}^2 - \frac{1}{s(r-1)} \left(\sum_{g=1}^s \sqrt{n_g} \sum_{j=1}^{r-1} y_{gj} \right)^2$$

$$(7.10) \quad = \sum \sum n_g y_{gj}^2 - \frac{1}{s(r-1)} \left[\sum_{g=1}^s \sqrt{n_g} (1 - y_{gr}) \right]^2,$$

$$(7.11) \quad \hat{u}'M\hat{u} = \hat{u}'M\hat{u} - \hat{y}'M\hat{y}$$

og følgelig

$$(7.12) \quad R^2 = \frac{\hat{y}'M\hat{y} - \frac{1}{s(r-1)} (i'M^{\frac{1}{2}}\hat{u})^2}{\hat{u}'M\hat{u} - \frac{1}{s(r-1)} (i'M^{\frac{1}{2}}\hat{u})^2}$$

7.1.2. Spesielt om binær regresjon

7.1.2.1 Teoretisk drøfting

Hvis den avhengige variable bare kan anta verdien 0 eller 1, er det åpenbart at restleddsvariansen vil være stor i forhold til det "gjennomsnittlige" (forventede) nivå på den avhengige variable. Forventningsverdien er lik sannsynligheten for at variabelen skal anta verdien 1. Spesielt når denne sannsynligheten er nær $\frac{1}{2}$, vil en 0-1-observasjon være et svært grovt mål på den underliggende ("sanne") sannsynlighet. I slike situasjoner er det rimelig å vente at andelen "uforklart varians", i regresjonen må bli relativt stor og føyningsmålet R^2 tilsvarende lavt. I dette avsnittet vil vi forsøke å si noe mer presist om dette forholdet, og å utlede hvor mye lavere R^2 vi kan vente oss i en situasjon med binær avhengig variabel.

Spørsmålet er av interesse fordi binær regresjon er en relativt vanlig metode i samfunnsvitenskapen, og fordi R^2 er det vanlige mål på "forklaringskraft" i en regresjonsmodell. Dersom forventningsverdien til den avhengige binærvariable virkelig kan uttrykkes som en lineær funksjon av et sett parametre, kan OLS i og for seg være en fornuftig metode. Vårt poeng her er imidlertid at føyningsmålet R^2 må tolkes med stor forsiktighet.

Tilfellet med binær avhengig variabel oppstår i vår modell som spesialtilfellet $r = 2$ og $n_g = 1$ ($g=1,2,\dots,s$), eller, om vi vil, dersom vi lar hvert individ telle som én observasjon, som i likning (4.57).

Determinasjonskoeffisienten for modellen (4.57) er gitt ved

$$(7.13) \quad R^{*2} = 1 - \frac{\hat{u}^{*'} \hat{u}^*}{\underline{y}^{*'} \underline{A} \underline{y}^*}.$$

Toppskriften * betyr at vi har å gjøre med en vektor eller matrise av individobservasjoner, slik vi skisserte i avsnitt 4.4. Her er

$$(7.14) \quad A = I - \frac{1}{n(r-1)} \underline{i} \underline{i}'$$

og \hat{u}^* er residualvektoren som framkommer ved OLS-estimering av modellen (4.57):

$$(7.15) \quad \hat{u}^* = \underline{y}^* - \hat{\underline{y}}^* = \underline{y}^* - Z^* \hat{\underline{\beta}} = \underline{y}^* - B^* \underline{y}^*$$

der

$$(7.16) \quad B^* = Z^* (Z^{*'} Z^*)^{-1} Z^{*'}.$$

B^* har egenskapen

$$(7.17) \quad B^{*2} = B^*.$$

Siden alle elementene i \underline{y}^* er enten 0 eller 1, vet vi at $y_{ij}^{*2} = y_{ij}^*$, og følgelig

$$(7.18) \quad \underline{y}^{*'} \underline{y}^* = \underline{i}' \underline{y}^*.$$

Vi har derfor

$$\begin{aligned} (7.19) \quad \hat{u}^{*'} \hat{u}^* &= (\underline{y}^* - B^* \underline{y}^*)' (\underline{y}^* - B^* \underline{y}^*) && \text{(iflg. (7.15))} \\ &= \underline{y}^{*'} \underline{y}^* - \underline{y}^{*'} B^* \underline{y}^* && \text{(iflg. (7.17))} \\ &= \underline{i}' \underline{y}^* - \hat{\underline{y}}^{*'} \underline{y}^* && \text{(iflg. (7.18))} \\ &= \sum_{g=1}^s \sum_{j=1}^{r-1} \sum_{i=q(g)+1}^{q(g+1)} (y_{ij}^* - \hat{y}_{ij}^{*2}) \\ &= \sum_{g=1}^s \sum_{j=1}^{r-1} n_g (y_{gj} - \hat{y}_{gj}^2) \\ &= \underline{i}' \underline{M} \underline{y} - \hat{\underline{y}}' \underline{M} \hat{\underline{y}}. \end{aligned}$$

Det nest siste likhetstegnet følger av (4.63) og av det faktum at OLS-estimering på modellene (4.57) og (4.75) gir samme estimator $\hat{\underline{\beta}}$, slik vi konstaterte i avsnitt 4.5. \hat{y}_{gj} er her de føyede verdier ved estimering på grupperte data, og vi har altså (jfr. (4.64))

$$(7.20) \quad \hat{y}_{ij}^* = \hat{y}_{gj} \text{ for } q(g)+1 \leq i \leq q(g+1).$$

Videre er

$$\begin{aligned}
 (7.21) \quad \underline{y}^*{}' A \underline{y}^* &= \underline{y}^*{}' \underline{y}^* - \frac{1}{n(r-1)} (\underline{i}' \underline{y}^*)^2 \\
 &= \underline{i}' \underline{y}^* - \frac{1}{n(r-1)} (\underline{i}' \underline{y}^*)^2 \\
 &= \underline{i}' \underline{M} \underline{y} - \frac{1}{n(r-1)} (\underline{i}' \underline{M} \underline{y})^2.
 \end{aligned}$$

Innsetting av (7.19) og (7.21) i (7.13) gir

$$(7.22) \quad R^{*2} = \frac{\hat{\underline{y}}' \underline{M} \hat{\underline{y}} - \frac{1}{n(r-1)} (\underline{i}' \underline{M} \hat{\underline{y}})^2}{\underline{i}' \underline{M} \underline{y} - \frac{1}{n(r-1)} (\underline{i}' \underline{M} \underline{y})^2}.$$

R^{*2} står altså for determinasjonskoeffisienten slik denne vil beregnes ved bruk av vanlig minste kvadraters metode (OLS) på individdata, jfr. modell (4.57). Til sammenlikning fant vi i forrige avsnitt (formel (7.12)) et uttrykk for determinasjonskoeffisienten (R^2) ved bruk av grupperte data til estimering av den samme parametervektor $\underline{\beta}$, jfr. modell (4.75). Til tross for at estimatet $\hat{\underline{\beta}}$ blir det samme ved begge metoder, blir R^{*2} forskjellig fra R^2 . La oss forsøke å se hvor stor forskjellen blir.

Det andre leddet i teller og nevner av (7.22) kan skrives

$$\begin{aligned}
 (7.23) \quad \frac{1}{n(r-1)} (\underline{i}' \underline{M} \underline{y})^2 &= \frac{1}{n(r-1)} \left[\sum_{g=1}^s n_g (1-y_{gr}) \right]^2 \\
 &= \frac{1}{s(r-1)} \left[\sum_{g=1}^s \sqrt{\frac{n_g}{n}} \cdot \sqrt{n_g} (1-y_{gr}) \right]^2
 \end{aligned}$$

der $\bar{n} = n/s$ er den gjennomsnittlige utvalgsstørrelsen for alle s grupper. Sammenliknet med det tilsvarende ledd i R^2 (jfr. (7.10)) veier vi altså hver gruppe med en faktor $\sqrt{n_g}/\bar{n}$. Subtrahendleddene i R^{*2} og R^2 er altså av samme størrelsesorden, men $\frac{1}{n(r-1)} (\underline{i}' \underline{M} \underline{y})^2$ vil være større enn $\frac{1}{s(r-1)} (\underline{i}' \underline{M} \underline{y})^2$ dersom andelene y_{gr} gjennomgående er små i de store utvalgsgruppene (n_g stor).

Første ledd i teller av R^{*2} og R^2 er like. Første ledd i nevner er imidlertid betydelig mindre for R^2 enn for R^{*2} , idet $0 \leq y_{gj} \leq 1$ og derfor

$$(7.24) \quad \underline{i}' \underline{M} \underline{y} = \sum_{g=1}^s \sum_{j=1}^{r-1} n_g y_{gj}^2 < \sum_{g=1}^s \sum_{j=1}^{r-1} n_g y_{gj} = \underline{i}' \underline{M} \underline{y}.$$

I praktisk talt ethvert tenkelig observasjonsmateriale vil vi derfor ha

$$(7.25) \quad R^{*2} < R^2.$$

Dersom (4.57) er den sanne modell, vet vi, ifølge de store tallers lov (A.20), at

$$(7.26) \quad \text{plim}_{n_g \rightarrow \infty} y_{gj} = \text{plim}_{n_g \rightarrow \infty} \hat{y}_{gj} = p_{gj} \quad (g=1,2,\dots,s; j=1,2,\dots,r),$$

altså at sannsynlighetsgrensen, når n_g vokser, for de observerte andelene y_{gj} , og også for de estimerte andelene \hat{y}_{gj} , er de sanne sannsynlighetene p_{gj} .

Vi har dessuten forutsatt (4.37) at

$$(7.27) \quad \lim_{n \rightarrow \infty} Q_n = \lim_{n \rightarrow \infty} \frac{1}{n} M = Q,$$

slik at

$$(7.28) \quad \text{plim}_{n \rightarrow \infty} \frac{1}{n} \underline{y}' M \underline{y} = \text{plim}_{n \rightarrow \infty} \underline{y}' Q_n \underline{y} = \underline{y}' Q \underline{y}$$

og tilsvarende

$$(7.29) \quad \text{plim}_{n \rightarrow \infty} \frac{1}{n} \hat{y}' M \hat{y} = \text{plim}_{n \rightarrow \infty} \frac{1}{n} \hat{y}' \hat{M} \hat{y} = \hat{y}' Q \hat{y}$$

$$(7.30) \quad \text{plim}_{n \rightarrow \infty} \left(\frac{1}{n} \underline{y}' M \underline{y} \right)^2 = (\underline{y}' Q \underline{y})^2$$

$$(7.31) \quad \lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} M^{\frac{1}{2}} = Q^{\frac{1}{2}}$$

$$(7.32) \quad \text{plim}_{n \rightarrow \infty} \frac{1}{n} (\underline{y}' M^{\frac{1}{2}} \underline{y})^2 = (\underline{y}' Q^{\frac{1}{2}} \underline{y})^2.$$

Samtlige resultater følger av setning A.1 og A.2 i Appendiks (avsnitt A.3.5).

Vi kan nå, ved å bruke de samme setninger, regne ut sannsynlighetsgrensene for de to føyingsmålene R^{*2} og R^2 :

$$(7.33) \quad \text{plim } R^{*2} = \frac{\text{plim}_{n \rightarrow \infty} \frac{1}{n} \hat{y}' M \hat{y} - \text{plim}_{n \rightarrow \infty} \left(\frac{1}{n} \underline{y}' M \underline{y} \right)^2}{\text{plim}_{n \rightarrow \infty} \frac{1}{n} \underline{y}' M \underline{y} - \text{plim}_{n \rightarrow \infty} \left(\frac{1}{n} \underline{y}' M \underline{y} \right)^2}$$

$$= \frac{\hat{y}' Q \hat{y} - (\underline{y}' Q \underline{y})^2}{\underline{y}' Q \underline{y} - (\underline{y}' Q \underline{y})^2} < 1.$$

$$(7.34) \quad \text{plim } R^2 = \frac{\text{plim}_{n \rightarrow \infty} \frac{1}{n} \hat{y}' M \hat{y} - \frac{1}{s} \text{plim}_{n \rightarrow \infty} \frac{1}{n} (\underline{y}' M^{\frac{1}{2}} \underline{y})^2}{\text{plim}_{n \rightarrow \infty} \frac{1}{n} \underline{y}' M \underline{y} - \frac{1}{s} \text{plim}_{n \rightarrow \infty} \frac{1}{n} (\underline{y}' M^{\frac{1}{2}} \underline{y})^2}$$

$$= \frac{\hat{y}' Q \hat{y} - \frac{1}{s} (\underline{y}' Q^{\frac{1}{2}} \underline{y})^2}{\underline{y}' Q \underline{y} - \frac{1}{s} (\underline{y}' Q^{\frac{1}{2}} \underline{y})^2} = 1.$$

Sannsynlighetsgrensen for R^2 er altså i vår modell lik 1. Det er et intuitivt helt opplagt resultat: Når utvalget innen hver gruppe g vokser mot uendelig, blir andelen y_{gj} et stadig mer presist estimat av den sanne sannsynligheten p_{gj} . Relasjon (7.27) innebærer at når n går mot uendelig, så går også alle n_g ($g=1,2,\dots,s$) mot uendelig. Dette fører til at restleddsvariansene $p_{gj}(1-p_{gj})/n_g$ går mot null, og vi får perfekt føyning forutsatt at sannsynlighetsmodellen er korrekt spesifisert.

Sannsynlighetsgrensen for R^{*2} er derimot i alminnelighet betydelig mindre enn 1. (Ulikheten (7.33) gjelder forutsatt at ikke alle de sanne sannsynlighetene p_{gj} er lik 0 eller 1.) Dette innser vi også lett intuitivt: økning i antall observasjoner fører ikke til at observasjonene y_{ij}^* nærmer seg sannsynlighetene p_{gj} , siden variablene y_{ij}^* fortsatt bare kan anta verdiene 0 og 1. Restleddsvariansen for individobservasjonene er derfor uavhengig av utvalgsstørrelsen.

plim R^{*2} avhenger av p og Q . Ved å regne ut for en del talleksempler vil vi undersøke hvordan plim R^{*2} avhenger av p . Dette kommer vi tilbake til mot slutten av avsnittet.

La oss først forsøke å utlede noe om egenskapene til R^{*2} og R^2 ved endelige (små) utvalg. I den forbindelse vil vi gjøre den forenklende forutsetningen

$$(7.35) \quad n_g = n/s \quad (g=1,2,\dots,s)$$

eller

$$(7.36) \quad M = nQ = \frac{n}{s} I$$

dvs. at utvalget fordeler seg helt jevnt på de s gruppene. (Vi ser fortsatt bare på binærtilfellet, dvs. $r = 2$ og følgelig $M = N$.)

Da har vi:

$$(7.37) \quad R^2 = \frac{\frac{n}{s} \hat{y}' \hat{y} - \frac{1}{s} \cdot \frac{n}{s} (\hat{y}' \mathbf{1})^2}{\frac{n}{s} \hat{y}' \hat{y} - \frac{1}{s} \cdot \frac{n}{s} (\hat{y}' \mathbf{1})^2}$$

$$= \frac{\hat{y}' \hat{y} - (\hat{y}' \mathbf{1})^2}{\hat{y}' \hat{y} - (\hat{y}' \mathbf{1})^2}$$

$$(7.38) \quad R^{*2} = \frac{\frac{n}{s} \hat{y}' \hat{y} - \frac{1}{n} \cdot \left(\frac{n}{s}\right)^2 \cdot (\hat{y}' \mathbf{1})^2}{\frac{n}{s} \hat{y}' \hat{y} - \frac{1}{n} \cdot \left(\frac{n}{s}\right)^2 \cdot (\hat{y}' \mathbf{1})^2}$$

$$= \frac{\hat{y}' \hat{y} - (\hat{y}' \mathbf{1})^2}{s \hat{y}' \hat{y} - (\hat{y}' \mathbf{1})^2}$$

og følgelig

$$(7.39) \quad \frac{R^{*2}}{R^2} = \frac{s \hat{y}' \hat{y} - (\hat{y}' \mathbf{1})^2}{s \hat{y}' \hat{y} - (\hat{y}' \mathbf{1})^2} < 1.$$

Kan vi si noe om forventningen til denne brøken?

Definer

$$(7.40) \quad \hat{\tau}_1 = s \hat{y}' \hat{y} - (\hat{y}' \mathbf{1})^2$$

$$(7.41) \quad \hat{\tau}_2 = s \hat{y}' \hat{y} - (\hat{y}' \mathbf{1})^2$$

slik at $R^{*2}/R^2 = \hat{\tau}_1/\hat{\tau}_2$. La τ_1 og τ_2 stå for forventningene til henholdsvis $\hat{\tau}_1$ og $\hat{\tau}_2$, og la oss dessuten definere

$$(7.42) \quad d_i = \frac{\hat{\tau}_i - \tau_i}{\tau_i} \quad (i=1,2).$$

d_i er således det relative avvik fra forventningsverdien til $\hat{\tau}_i$, og vi forutsetter $d_i < 1$ ($i=1,2$).
Da kan vi skrive

$$(7.43) \quad \begin{aligned} E(\hat{\tau}_1/\hat{\tau}_2) &= E\left[\frac{\tau_1(d_1+1)}{\tau_2(d_2+1)}\right] \\ &= \frac{\tau_1}{\tau_2} E\left(\frac{d_1+1}{d_2+1}\right) = \frac{\tau_1}{\tau_2} + \frac{\tau_1}{\tau_2} E\left(\frac{d_1-d_2}{d_2+1}\right) \\ &= \frac{\tau_1}{\tau_2} \left[1 + E\left\{ (d_1-d_2)(1-d_2+d_2^2-d_2^3+\dots) \right\} \right] \\ &\approx \frac{\tau_1}{\tau_2} \left[1 + E(d_1-d_2-d_1d_2+d_2^2) \right], \end{aligned}$$

der vi har Taylor-utviklet $\frac{1}{d_2+1}$ og sett bort fra 3.-ordens-ledd og høyere. Nå er

$$(7.44) \quad E(d_1) = E(d_2) = 0$$

$$(7.45) \quad E(d_1d_2) = \frac{1}{\tau_1\tau_2} \text{cov}(\hat{\tau}_1, \hat{\tau}_2)$$

$$(7.46) \quad E(d_2^2) = \frac{1}{\tau_2^2} \text{var}(\hat{\tau}_2),$$

og vi finner

$$(7.47) \quad E(\hat{\tau}_1/\hat{\tau}_2) \approx \frac{\tau_1}{\tau_2} + \frac{\tau_1}{\tau_2} \left[\frac{1}{\tau_2^2} \text{var}(\hat{\tau}_2) - \frac{1}{\tau_1\tau_2} \text{cov}(\hat{\tau}_1, \hat{\tau}_2) \right].$$

Utregning av $\text{var}(\hat{\tau}_2)$ og $\text{cov}(\hat{\tau}_1, \hat{\tau}_2)$ leder til en del nokså uhåndterlige matematiske uttrykk, og vi har derfor valgt den enklere løsning å se på brøken τ_1/τ_2 , altså på forholdet mellom forventningsverdiene til teller og nevner i (7.39). τ_1/τ_2 vil være forskjellig fra forventningsverdien til R^2/R^2 i den grad $\frac{1}{\tau_2} \text{var}(\hat{\tau}_2)$ avviker fra $\frac{1}{\tau_1} \text{cov}(\hat{\tau}_1, \hat{\tau}_2)$. I vårt tilfelle kan vi håpe at dette avviket ikke er svært stort, idet både $\hat{\tau}_1$ og $\hat{\tau}_2$ varierer i samme retning som funksjoner av y_{g1} ($g=1,2,\dots,s$). Således vet vi i det minste at $\text{cov}(\hat{\tau}_1, \hat{\tau}_2)$ har positivt fortegn.

Nå er

$$(7.48) \quad \begin{aligned} E(\underline{y}'\underline{y}) &= E\left(\sum_{g=1}^s y_{g1}^2\right) = \sum_{g=1}^s E(y_{g1}^2) \\ &= \sum_{g=1}^s \text{var}(y_{g1}) + \sum_{g=1}^s \left[E(y_{g1}) \right]^2 \\ &= \sum_{g=1}^s \frac{s}{n} p_{g1}(1-p_{g1}) + \sum_{g=1}^s p_{g1}^2 \\ &= \frac{s}{n} \sum_{g=1}^s p_{g1} + \frac{n-s}{n} \sum_{g=1}^s p_{g1}^2 \end{aligned}$$

$$(7.49) \quad E(\bar{y}) = E\left(\sum_{g=1}^s y_{g1}\right) = \sum_{g=1}^s p_{g1}$$

$$(7.50) \quad E\left[\left(\sum_{g=1}^s y_{g1}\right)^2\right] = E\left[\left(\sum_{g=1}^s y_{g1}\right)^2\right]$$

$$= \sum_{g=1}^s \sum_{h=1}^s E(y_{g1}y_{h1})$$

$$= \sum_{g=1}^s \sum_{h=1}^s E(y_{g1}) \cdot E(y_{h1}) + \sum_{g=1}^s \text{var}(y_{g1})$$

$$= \sum_{g=1}^s \sum_{h=1}^s p_{g1}p_{h1} + \sum_{g=1}^s \frac{s}{n} p_{g1}(1-p_{g1})$$

$$= \left(\sum_{g=1}^s p_{g1}\right)^2 + \frac{s}{n} \sum_{g=1}^s p_{g1} - \frac{s}{n} \sum_{g=1}^s p_{g1}^2.$$

Ved å kombinere disse uttrykkene med (7.40) og (7.41) får vi

$$(7.51) \quad \tau_1 = \frac{s(s-1)}{n} \sum_{g=1}^s p_{g1} - \left(\sum_{g=1}^s p_{g1}\right)^2 + \frac{s(n-s+1)}{n} \left(\sum_{g=1}^s p_{g1}^2\right)$$

$$(7.52) \quad \tau_2 = \frac{s(n-1)}{n} \sum_{g=1}^s p_{g1} - \left(\sum_{g=1}^s p_{g1}\right)^2 + \frac{s}{n} \sum_{g=1}^s p_{g1}^2.$$

Forholdstallet τ_1/τ_2 avhenger av s , n og p_{11}, \dots, p_{s1} . Definer

$$(7.53) \quad \mu = \frac{1}{s} \sum_{g=1}^s p_{g1}$$

$$(7.54) \quad \sigma^2 = \frac{1}{s} \sum_{g=1}^s p_{g1}^2 - \mu^2$$

og

$$(7.55) \quad \omega = \mu(1-\mu).$$

Da kan vi skrive

$$(7.56) \quad \frac{\tau_1}{\tau_2} = \frac{(n-s+1)\sigma^2 + (s-1)\omega}{\sigma^2 + (n-1)\omega}$$

og følgelig

$$(7.57) \quad \lim_{n \rightarrow \infty} (\tau_1/\tau_2) = \sigma^2/\omega.$$

Forholdstallet τ_1/τ_2 avhenger av p_{11}, \dots, p_{s1} bare gjennom σ^2 og ω . σ^2 uttrykker spredningen, mens ω er større jo mindre gjennomsnittet av p-ene avviker fra 0,5.

Ved partiell derivasjon av (7.56) finner vi

$$(7.58) \quad \frac{\partial(\tau_1/\tau_2)}{\partial(\sigma^2)} = \frac{n(n-s)\omega}{[\sigma^2+(n-1)\omega]^2} > 0,$$

$$(7.59) \quad \frac{\partial(\tau_1/\tau_2)}{\partial\omega} = -\frac{n(n-s)\sigma^2}{[\sigma^2+(n-1)\omega]^2} < 0$$

og

$$(7.60) \quad \frac{\partial^2(\tau_1/\tau_2)}{\partial\omega\partial(\sigma^2)} = \frac{n(n-s)[\sigma^2-(n-1)\omega]}{[\sigma^2+(n-1)\omega]^3} < 0,$$

siden $0 < \sigma^2 < \omega < 0,25$. Ulikhetene gjelder forutsatt at ikke alle p-er er lik enten 0 eller 1.

Vi kan slutte at, for gitte verdier av n og s, blir R^{*2} mindre i forhold til R^2 jo mindre spredning det er på de sanne sannsynlighetene, og jo mer midt i $[0,1]$ - intervallet gjennomsnittet ligger.

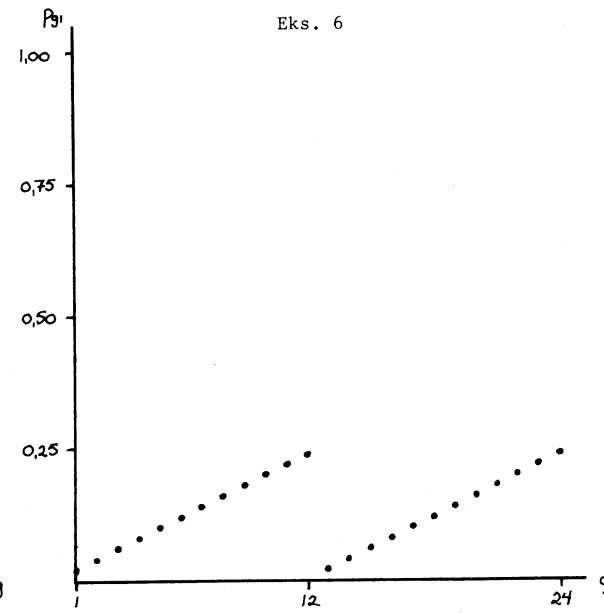
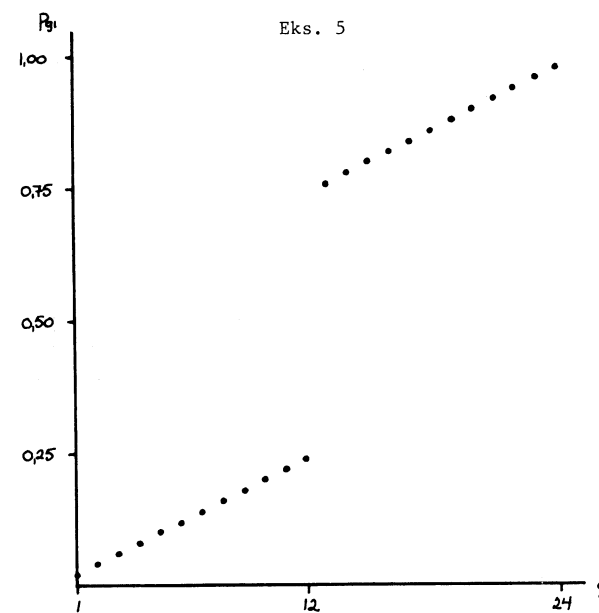
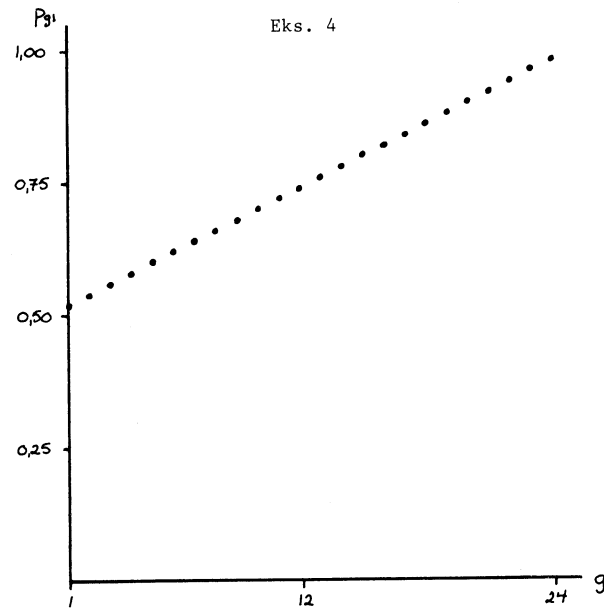
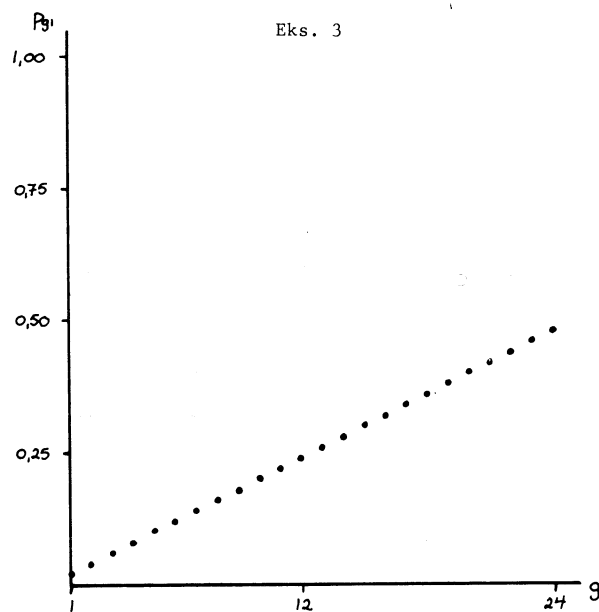
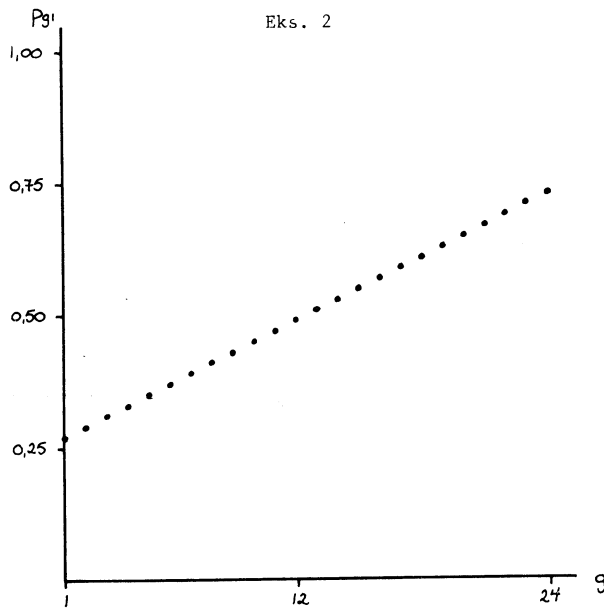
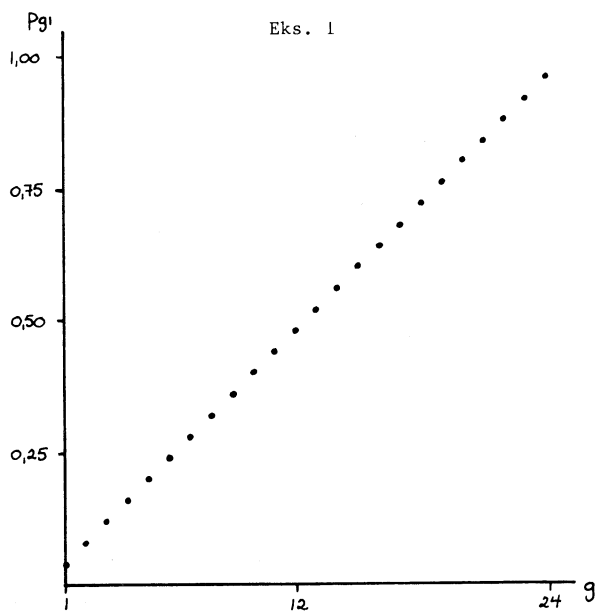
7.1.2.2 Regneeksempler

Vi har konstruert en del talleksempler for ytterligere å illustrere dette forholdet. Eksemplene framgår av tabell 7.1 og er dessuten gjengitt i form av figur 7.1.

Tabell 7.1. Talleksempler for sannsynlighetene p_{g1} ($g=1,2,\dots,s$) og formler for utregning av ω , σ^2 og $\text{plim } R^{*2}$ *Numerical examples for the true probabilities p_{g1} ($g=1,2,\dots,s$) and formulae for calculating ω , σ^2 , and $\text{plim } R^{*2}$*

Eksempel nr. Example no.	g	p_{g1}	ω	σ^2	$\text{plim } R^{*2}$
1	1,2,...,s	$\frac{g}{s+1}$	$\frac{1}{4}$	$\frac{s-1}{12(s+1)}$	$\frac{s-1}{3(s+1)}$
2	1,2,...,s	$\frac{g}{2(s+1)} + \frac{1}{4}$	$\frac{1}{4}$	$\frac{s-1}{48(s+1)}$	$\frac{s-1}{12(s+1)}$
3	1,2,...,s	$\frac{g}{2(s+1)}$	$\frac{3}{16}$	$\frac{s-1}{48(s+1)}$	$\frac{s-1}{9(s+1)}$
4	1,2,...,s	$\frac{g}{2(s+1)} + \frac{1}{2}$	$\frac{3}{16}$	$\frac{s-1}{48(s+1)}$	$\frac{s-1}{9(s+1)}$
5	$\left\{ \begin{array}{l} 1,2,\dots,\frac{s}{2} \\ \frac{s}{2}+1,\dots,s \end{array} \right.$	$\left. \begin{array}{l} \frac{g}{2(s+1)} \\ \frac{g}{2(s+1)} + \frac{1}{2} \end{array} \right\}$	$\frac{1}{4}$	$\frac{7s+2}{48(s+1)}$	$\frac{7s+2}{12(s+1)}$
6	$\left\{ \begin{array}{l} 1,2,\dots,\frac{s}{2} \\ \frac{s}{2}+1,\dots,s \end{array} \right.$	$\left. \begin{array}{l} \frac{g}{2(s+1)} \\ \frac{g-s/2}{2(s+1)} \end{array} \right\}$	$\frac{(s+2)(7s+6)}{64(s+1)^2}$	$\frac{(s+2)(s-2)}{192(s+1)^2}$	$\frac{s-2}{3(7s+6)}$

Figur 7.1. Sannsynlighetene i eks. 1-6, opptegnet for $s = 24$ The probabilities in ex. 1-6, illustrated for $s = 24$



I eks. 1 tenker vi oss at sannsynlighetene er jevnt fordelt fra en minste verdi $p_{11} = \frac{1}{s+1}$ til en høyeste verdi $p_{s1} = 1 - \frac{1}{s+1}$. Det er med andre ord stor spredning i de sanne sannsynlighetene, og sannsynlighetene fordeler seg symmetrisk rundt 0,5.

I eks. 2 er sannsynlighetene fortsatt symmetriske, men med langt mindre spredning - de ligger alle mellom 0,25 og 0,75.

I eks. 3 fordeler sannsynlighetene seg jevnt mellom 0 til 0,5, mens de i eks. 4 ligger mellom 0,5 og 1. Begge eksemplene 3 og 4 er altså "skjeve", men til forskjellig side.

I eks. 5 ligger halvparten av sannsynlighetene mellom 0 og 0,25 og den andre halvparten mellom 0,75 og 1. I eks. 6 ligger den første halvparten som i eks. 5, mens den andre halvparten her antas å være lik den første. Begge eksempler har altså sannsynligheter som er like "ekstreme", i den forstand at de ligger like langt fra 0,5. Eks. 5 er imidlertid symmetrisk og med stor spredning, mens eks. 6 er "skjevt" og har liten spredning.

Ved å bruke de velkjente formlene $\sum_{g=1}^s g = s(s+1)/2$ og $\sum_{g=1}^s s^2 = s(s+1)(2s+1)/6$ kan vi finne enkle aritmetiske uttrykk for ω og σ^2 i hvert av de 6 eksemplene. Formlene framgår av tabell 7.1. Ved hjelp av en programmerbar lommekalkulator har vi regnet ut τ_1/τ_2 for en del alternative valg av n og s . Resultatene er gjengitt i tabellene 7.2 - 7.6. Lengst til høyre i tabellene har vi en kolonne for plim $R^{*2} = \text{plim}(R^{*2}/R^2)$. plim R^{*2} faller, som seg hør og bør, sammen med $\lim_{n \rightarrow 0} (\tau_1/\tau_2)$. Også plim R^{*2} blir i talleksemplene en svært enkel funksjon av s .

Eksemplene 3 og 4 er i praksis like. Innsetting av $(1-p_{g1})$ istedenfor p_{g1} i uttrykkene (7.51) og (7.52) gir samme resultat. Det spiller altså ingen rolle hvilket av de to utfall vi velger å kalle "1" og hvilket vi kaller "0".

Verdien av τ_1/τ_2 viser seg å variere sterkt med n , og i langt mindre grad med s . Dette er et resultat vi måtte vente: presisjonsgevinsten ved å bruke gjennomsnittet y_{gj} istedenfor enkeltobservasjonene y_{ij}^* må selvsagt avhenge sterkt av hvor mange observasjoner vi kan basere hvert gjennomsnitt på.

Tabell 7.2. Eksempel 1. Verdien av τ_1/τ_2 og av plim R^{*2} , for utvalgte verdier av s og n *Example 1.*
Value of τ_1/τ_2 and of plim R^{*2} , for selected values of s and n

s	τ_1/τ_2							plim R^{*2}
	$\frac{n}{s} = 1$	$\frac{n}{s} = 2$	$\frac{n}{s} = 5$	$\frac{n}{s} = 10$	$\frac{n}{s} = 100$	$\frac{n}{s} = 500$	$\frac{n}{s} = 1\ 000$	
5	1	0.578	0.358	0.289	0.229	0.224	0.223	0.222
10	1	0.623	0.410	0.341	0.279	0.274	0.273	0.273
50	1	0.658	0.455	0.387	0.327	0.322	0.321	0.320
100	1	0.662	0.461	0.394	0.333	0.328	0.327	0.327
500	1	0.666	0.465	0.399	0.339	0.333	0.333	0.332

Tabell 7.3. Eksempel 2. Verdien av τ_1/τ_2 og av $\text{plim } R^{*2}$, for utvalgte verdier av s og n *Example 2.*
*Value of τ_1/τ_2 and of $\text{plim } R^{*2}$, for selected values of s and n*

s	τ_1/τ_2							plim R^{*2}
	$\frac{n}{s} = 1$	$\frac{n}{s} = 2$	$\frac{n}{s} = 5$	$\frac{n}{s} = 10$	$\frac{n}{s} = 100$	$\frac{n}{s} = 500$	$\frac{n}{s} = 1\ 000$	
5	1	0.479	0.215	0.134	0.063	0.057	0.056	0.056
10	1	0.511	0.240	0.153	0.077	0.070	0.069	0.068
50	1	0.536	0.261	0.171	0.089	0.082	0.081	0.080
100	1	0.539	0.264	0.173	0.091	0.084	0.083	0.082
500	1	0.541	0.266	0.175	0.092	0.085	0.084	0.084

Tabell 7.4. Eksempel 3 og 4. Verdien av τ_1/τ_2 og av $\text{plim } R^{*2}$, for utvalgte verdier av s og n *Examples 3 and 4.*
*Value of τ_1/τ_2 and of $\text{plim } R^{*2}$, for selected values of s and n*

s	τ_1/τ_2							plim R^{*2}
	$\frac{n}{s} = 1$	$\frac{n}{s} = 2$	$\frac{n}{s} = 5$	$\frac{n}{s} = 10$	$\frac{n}{s} = 100$	$\frac{n}{s} = 500$	$\frac{n}{s} = 1\ 000$	
5	1	0.490	0.231	0.151	0.082	0.076	0.075	0.074
10	1	0.524	0.259	0.174	0.099	0.093	0.092	0.091
50	1	0.549	0.283	0.195	0.116	0.109	0.108	0.107
100	1	0.552	0.286	0.197	0.118	0.111	0.110	0.109
500	1	0.555	0.288	0.199	0.120	0.112	0.112	0.111

Tabell 7.5. Eksempel 5. Verdien av τ_1/τ_2 og av $\text{plim } R^{*2}$, for utvalgte verdier av s og n *Example 5.*
*Value of τ_1/τ_2 and of $\text{plim } R^{*2}$, for selected values of s and n*

s	τ_1/τ_2							plim R^{*2}
	$\frac{n}{s} = 1$	$\frac{n}{s} = 2$	$\frac{n}{s} = 5$	$\frac{n}{s} = 10$	$\frac{n}{s} = 100$	$\frac{n}{s} = 500$	$\frac{n}{s} = 1\ 000$	
5
10	1	0.767	0.633	0.589	0.550	0.546	0.546	0.545
50	1	0.787	0.660	0.617	0.579	0.576	0.576	0.575
100	1	0.789	0.663	0.621	0.583	0.580	0.580	0.579
500	1	0.791	0.666	0.624	0.587	0.583	0.583	0.583

Tabell 7.6. Eksempel 6. Verdien av τ_1/τ_2 og av $\text{plim } R^{*2}$, for utvalgte verdier av s og n *Example 6.*
*Value of τ_1/τ_2 and of $\text{plim } R^{*2}$, for selected values of s and n*

s	τ_1/τ_2							plim R^{*2}
	$\frac{n}{s} = 1$	$\frac{n}{s} = 2$	$\frac{n}{s} = 5$	$\frac{n}{s} = 10$	$\frac{n}{s} = 100$	$\frac{n}{s} = 500$	$\frac{n}{s} = 1\ 000$	
5
10	1	0.493	0.213	0.123	0.044	0.037	0.036	0.035
50	1	0.518	0.233	0.139	0.054	0.047	0.046	0.045
100	1	0.521	0.236	0.141	0.056	0.048	0.047	0.046
500	1	0.523	0.238	0.142	0.057	0.049	0.048	0.047

I spesialtilfellet $n_g = \frac{n}{s} = 1$ blir selvsagt τ_1/τ_2 lik 1 - her består de to metodene av nøyaktig samme prosedyre.

Allerede når vi har $\frac{n}{s} = 2$ observasjoner for hver celle er imidlertid forskjellen mellom τ_1 og τ_2 drastisk - forholdstallet varierer fra under 50 til nærmere 80 prosent i våre eksempler. Når n går mot uendelig finner vi at R^{*2} i flere tilfelle blir bare 5-10 prosent av R^2 , som går mot 1.

Som ventet blir R^{*2} aller lavest for eksemplene 2, 3, 4 og 6. Her er spredningen liten og/eller gjennomsnittet "skjevt", og en kan ikke vente determinasjonskoeffisienter på over 0,10.

For enkelhets skyld har vi regnet på et antall eksempler der det er flere enn én observasjon for hver kombinasjon av verdier på de uavhengige variable. Konklusjonen om gjennomgående lav R^{*2} ved binær regresjon kan imidlertid ikke avhenge av om de uavhengige variable er av denne spesielle typen. Det samme vil gjelde dersom vi har kontinuerlige forklaringsvariable, selv om vi i dette tilfellet ikke har noe naturlig alternativt føyningsmål (R^2).

En konklusjon en kan trekke av dette er at, som mål på sannsynlighetsmodellens "godhet" eller forklaringskraft, må R^{*2} tolkes med stor forsiktighet. I beste fall kan R^{*2} , for et gitt observasjonsmateriale og gitte p_{11}, p_{21}, \dots , si noe om hvilket valg av uavhengige variable som gir best tilpasning. Sammenlikning mellom R^{*2} for ulike datasett og ulike "sanne" sannsynligheter gir imidlertid liten mening. Verdien på R^{*2} er i første rekke bestemt av helt andre forhold enn hvorvidt regresjonslikningen er korrekt spesifisert.

7.2 Kji-kvadrat-mål

7.2.1 Pearson kji-kvadrat (χ_p^2)

Det vanligste mål for føyning i en kontingenstabell er Pearson kji-kvadrat-observatoren, som ble definert i avsnitt 4.7. Anvendt på en generell kontingenstabell kan denne observatoren skrives

$$(7.61) \quad \chi_p^2 = \sum (o-e)^2/e$$

der o står for det observerte antall tilfeller i en celle, og e for den estimerte forventningen av o , dvs. den føyede verdi i cellen. Summasjonen går over alle celler i tabellen.

χ_p^2 er som nevnt asymptotisk kji-kvadrat-fordelt under nullhypotesen at sannsynlighetsmodellen er korrekt spesifisert. For at kji-kvadrat-tilnærmingen skal være brukbar i endelige utvalg, er det vanlig å nytte en tommelfingerregel om at e ikke for noen celle bør være mindre enn 5. Enkelte statistikere er mindre strenge og setter skillet ved $e = 3$, eller enda lavere.

7.2.2 Neyman kji-kvadrat (χ_N^2)

Neyman kji-kvadrat-observatoren ble likeledes definert i avsnitt 4.7. Denne observatoren er mest kjent under navnet modifisert Pearson kji-kvadrat. Modifikasjonen består i å erstatte e i nevneren med o , altså

$$(7.62) \quad \chi_N^2 = \sum(o-e)^2/o.$$

I avsnitt 4.7 viste vi at når 2SGLS nyttes til estimering av e under den lineære modell, så leder dette pr. definisjon til minimering av χ_N^2 . χ_N^2 har samme asymptotiske egenskaper som χ_p^2 , og en nytter gjerne de samme tommelfingerregler for når kji-kvadrat-tilnærmingen kan brukes til testing i endelige utvalg.

7.2.3 Kji-kvadrat-observatoren basert på sannsynlighetsknoten (G^2)

Størrelsen

$$(7.63) \quad G^2 = 2\sum o \cdot \log(o/e)$$

ble definert i avsnitt 5.4, der vi også slo fast at dette er den såkalte "kji-kvadratobservatoren basert på sannsynlighetsknoten", som minimeres ved sannsynlighetsmaksimeringsmetoden anvendt på den multinomiske logit-modell eller den symmetriske log-lineære modell. Vi innser lett at (7.63) = (5.53) = (5.56).

G^2 er, i likhet med χ_p^2 og χ_N^2 , asymptotisk kji-kvadrat-fordelt, men er ved endelige utvalg antakelig mindre sensitiv overfor tilfeldige variasjoner i "små" celler enn χ_p^2 og χ_N^2 .

7.3 Feilklassifiseringsindeksen (C)

Et intuitivt rimelig føyningsmål får vi ved å telle opp hvor mange individer som blir klassifisert feil gjennom estimeringsmetoden, dvs. at vi for hver celle teller opp

$$(7.64) \quad |o-e| = n_g |\hat{u}_{gj}| = n_g |y_{gj} - \hat{y}_{gj}|,$$

og summerer. Metoden består m.a.o. i summasjon av de absolutte avvik eller residualer. Siden $\sum_{j=1}^r \hat{u}_{gj} = 0$ for alle g innebærer metoden at hver "feilklassifisering" blir talt opp to ganger. Vi foreslår derfor å bruke føyningsmålet

$$(7.65) \quad C = \frac{1}{2} \sum_{g=1}^s \sum_{j=1}^r n_g |y_{gj} - \hat{y}_{gj}| = \sum_{\{\hat{u}_{gj} > 0\}} n_g (y_{gj} - \hat{y}_{gj}).$$

Denne observatoren kan gis en rekke nyttige tolkninger. For det første er den identisk med et føyningsmål foreslått av Domencich og McFadden (1975, s.124-125), der C er gitt en tolkning som den totale kostnaden ved feilklassifisering når feilklassifisering av hvert individ teller (koster) like meget (1 kr). Observatoren $\frac{100}{n} C$ kan tolkes som prosentandelen av det totale materialet som feilklassifiseres. $\frac{2}{rn} C$ er (det veide) gjennomsnittet av alle absolutte avvik $|\hat{u}_{gj}|$. Målet $100 \cdot \frac{2}{rn} \cdot C$ er det (veid) gjennomsnittlige avviket mellom observerte og føyede andeler, målt i prosentpoeng.

C har ingen kjent fordeling og er motivert ut fra ønsket om å ha et føyningsmål som er mest mulig nøytralt i forhold til de ulike sannsynlighetsmodeller og estimeringsmetoder.

7.4 Sammenfatning om føyningsmål

Vi har til rådighet en rekke ulike føyningsmål for kontingenstabeller og modeller med kvalitative avhengige variable. Felles for de fleste av de målene vi har satt opp er at hvert mål minimeres gjennom bruk av én bestemt estimeringsmetode på en gitt sannsynlighetsmodell. R^2 minimeres av OLS, dersom vi har lineær sannsynlighetsmodell. χ_N^2 minimeres av 2SGLS, gitt lineær sannsynlighetsmodell. G^2 minimeres av ML, gitt log-lineær sannsynlighetsmodell.

Ingen av disse føyningsmålene er altså nøytrale i forhold til estimeringsmetode og sannsynlighetsmodell. Bruk av målet R^2 for sammenlikning mellom OLS og 2SGLS i den lineære sannsynlighetsmodell er f.eks. "urettferdig" overfor 2SGLS, fordi OLS nettopp går ut på minimering av R^2 .

For å få til en empirisk sammenlikning mellom de tre metodene har vi derfor, for hver metode, regnet ut alle de tre føyningsmålene R^2 , χ_N^2 og G^2 (se kapittel 9). Utrekningene er gjennomført under to alternative forutsetninger om samspillseffekter. I tillegg har vi beregnet føyningsobservatoren C, som kan begrunnes intuitivt og uavhengig av hvilken sannsynlighetsmodell eller estimeringsmetode som anses "riktig".

8. OPPLÈGG FOR DEN EMPIRISKE ANALYSEN

8.1 Omfang

Vi har nyttet et datamateriale fra Statistisk Sentralbyrås arbeidskraftundersøkelser (AKU) for 1976 til utprøving av følgende statistiske modeller/estimeringsmetoder: (i) vanlig minste kvadraters metode (OLS) i den lineære modell, (ii) to-trinns generaliserte minste kvadraters metode (2SGLS) i den lineære modell, og (iii) sannsynlighetsmaksimeringsmetoden (ML) i den log-lineære modell/multinomske logit-modell.

8.2 Datamaterialet

Datamaterialet er en kontingenstabell med 9 523 observasjoner, bestående av menn mellom 16 og 74 år som har besvart AKU-skjemaet for et eller annet kvartal i 1976. Vi bruker imidlertid ikke hele AKU-materialet: i første og tredje kvartal har vi trukket ut personer med fødselsdato 1-15, og i annet og fjerde kvartal personer med fødselsdato 16-31. Antall observasjoner blir dermed halvert og vi oppnår at ingen person er med i utvalget mer enn én gang. (AKU har roterende utvalg, jfr. Statistisk Sentralbyrå (1978).) Dermed kan vi, uten særlig stor feil, anta at alle observasjoner i utvalget er ukorrelet med hverandre.

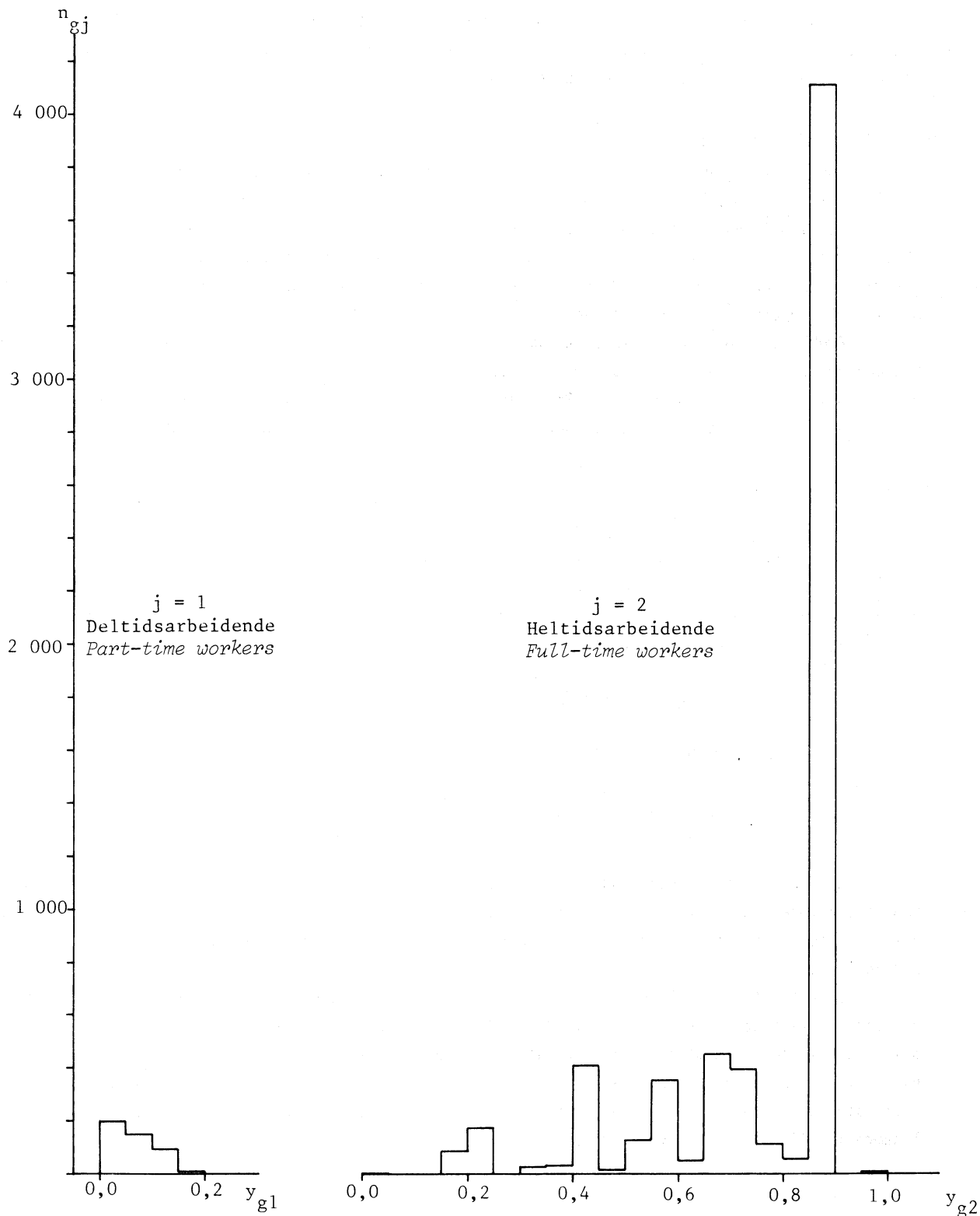
Kontingenstabellen er gjengitt som tabell 8.1. Den har 4 grupperingsveier (variable) og 135 celler.

Den avhengige variable har vi kalt arbeidstid (T). Den har 3 nivåer: 0 timer ("ikke sysselsatt i inntektsgivende arbeid"), 1-29 timer ("sysselsatt på deltid"), og minst 30 timer ("sysselsatt på heltid"). Som uavhengige variable har vi sivilstand (S) (3 nivåer), utdanningsnivå (U) (3 nivåer) og alder (A) (5 nivåer). Alle variable er kvalitative, dvs. målt på (eller redusert til) nominal- eller ordinalnivå.

Tabell 8.1. Datamaterialet. Menn 16-74 år, etter sivilstand, utdanningsnivå, alder og arbeidstid. Absolutte tall (n_{gj}) og prosent ($100 \cdot y_{gj}$) *The data set. Men 16-74 years old, by marital status, level of education, age, and weekly number of hours worked. Absolute numbers (n_{gj}) and per cent ($100 \cdot y_{gj}$)*

	Arbeidstimer pr. uke <i>Hours worked per week</i>							
	I alt <i>Total</i>		0		1-29		30 +	
	n_g	$100 \cdot \frac{\sum y_{gj}}{j}$	n_{g3}	$100 \cdot y_{g3}$	n_{g1}	$100 \cdot y_{g1}$	n_{g2}	$100 \cdot y_{g2}$
Ugifte <i>Unmarried</i>								
Høyst 9 års utdanning <i>At most 9 years of schooling</i>								
16-19 år <i>years</i>	525	100	349	66	58	11	118	22
20-24 " "	184	100	73	40	10	5	101	55
25-59 " "	294	100	81	28	14	5	199	68
60-66 " "	74	100	28	38	5	7	41	55
67-74 " "	86	100	66	77	7	8	13	15
10-12 års utdanning <i>10-12 years of schooling</i>								
16-19 år	289	100	141	49	23	8	125	43
20-24 "	620	100	324	52	22	4	274	44
25-59 "	407	100	89	22	14	3	304	75
60-66 "	21	100	10	48	-	-	11	52
67-74 "	24	100	15	63	1	4	8	33
Minst 13 års utdanning <i>At least 13 years of schooling</i>								
16-19 år	1	100	-	-	-	-	1	100
20-24 "	74	100	41	55	4	5	29	39
25-59 "	120	100	41	34	9	8	70	58
60-66 "	4	100	1	25	-	-	3	75
67-74 "	3	100	2	67	1	33	-	-
Gifte <i>Married</i>								
Høyst 9 års utdanning								
16-19 år	10	100	6	60	2	20	2	20
20-24 "	44	100	8	18	-	-	36	82
25-59 "	1 520	100	179	12	43	3	1 298	85
60-66 "	420	100	152	36	25	6	243	58
67-74 "	379	100	292	77	25	7	62	16
10-12 års utdanning								
16-19 år	3	100	-	-	1	33	2	67
20-24 "	119	100	25	21	5	4	89	75
25-59 "	2 400	100	230	10	56	2	2 114	88
60-66 "	367	100	89	24	24	7	254	69
67-74 "	256	100	172	67	29	11	55	21
Minst 13 års utdanning								
16-19 år	-	-	-	-	-	-	-	-
20-24 "	18	100	9	50	1	6	8	44
25-59 "	794	100	56	7	38	5	700	88
60-66 "	66	100	11	17	4	6	51	77
67-74 "	52	100	29	56	7	13	16	31
Før gifte <i>Previously married</i>								
Høyst 9 års utdanning								
16-19 år	2	100	1	50	1	50	-	-
20-24 "	2	100	1	50	-	-	1	50
25-59 "	76	100	24	32	3	4	49	64
60-66 "	35	100	16	46	3	9	16	46
67-74 "	53	100	47	89	4	8	2	4
10-12 års utdanning								
16-19 år	1	100	-	-	-	-	1	100
20-24 "	3	100	-	-	-	-	3	100
25-59 "	80	100	14	18	2	3	64	80
60-66 "	29	100	12	41	2	7	15	52
67-74 "	38	100	26	68	6	16	6	16
Minst 13 års utdanning								
16-19 år	-	-	-	-	-	-	-	-
20-24 "	1	100	-	-	-	-	1	100
25-59 "	20	100	1	5	2	10	17	85
60-66 "	6	100	1	17	-	-	5	83
67-74 "	3	100	2	67	-	-	1	33

Figur 8.1. Histogrammer som viser hvordan de deltids- og heltidsarbeidende personene i utvalget fordeler seg etter den gjennomsnittlige deltids-, henh. heltidsandel blant personer i samme alders, utdannings- og sivilstandsgruppe.
Histograms showing the distribution of part-time and full-time workers in the sample, by the proportion of part-time and full-time workers, respectively, among respondents with identical values on the independent variables.



Observasjonene fordeler seg svært ujevnt på de 135 cellene. De observerte relative frekvensene y_{gj} fordeler seg også svært ujevnt over $[0,1]$ -intervallet. Dette er forsøkt gjengitt i figur 8.1. Langs x-aksen måler vi andelen deltids- eller heltidsarbeidende (y_{gj} , $j=1,2$) i hver gruppe, i intervaller på 5 prosent. Langs y-aksen har vi avsatt hvor mange deltids- eller heltidsarbeidende (n_{gj}) som tilhører grupper der y_{gj} faller i et bestemt intervall. Figuren fordeler således de deltids- eller heltidsarbeidende i utvalget etter den gjennomsnittlige andel deltids- eller heltidsarbeidende, henholdsvis, blant personer med samme verdi på de tre uavhengige variable alder, utdanning og sivilstand.

Vi ser at nærmere halvparten av personene i utvalget tilhører grupper der mellom 85 og 90 prosent arbeider heltid. Så å si ingen av gruppene har heltids-andeler på over 90 eller under 15 prosent. Derimot ligger deltids-andelen i nesten alle grupper under 15 prosent.

8.3 Problemet med tomme celler

Som det framgår av tabell 8.1, inneholder datamaterialet et ganske stort antall "tomme" celler, dvs. nuller. Av 135 celler er 23 tomme.

Teorien i avsnitt 4 og 5 er basert på forutsetningen om at alle de sanne sannsynlighetene p_{gj} er ekte positive og ekte mindre enn 1. Estimeringsmetodene er dessuten basert på forutsetningen om at alle observerte andeler y_{gj} også har denne egenskap. For å kunne anvende vårt datamateriale til estimering må vi derfor gjøre visse tillegninger.

Problemet med ingen eller bare svært få observasjoner i enkelte celler er nokså alminnelig. Det dukker opp i de fleste fingrupperede datamaterialer.

Haldane (1955) og Anscombe (1956) har foreslått å øke antallet observasjoner i hver celle med et tillegg $\delta = \frac{1}{2}$. Anscombe viser at dette ofte leder til mindre skjevhet i estimeringen av "log-oddsen"

$$(8.1) \quad \log\left(\frac{p_{g1}}{1-p_{g1}}\right) = x'_{g1}\beta_1$$

for den binomiske logit-modell. Samme prosedyre anbefales av Cox (1970). Hitchcock (1962) foreslår $\delta = \frac{1}{4}$ og viser at dette som regel gir mindre skjevhet for parameterestimatorene $\hat{\beta}_1$ enn $\delta = \frac{1}{2}$. Andre metoder har også vært foreslått. Berkson (1955) går inn for den såkalte "2n-regelen", som i vårt tilfelle ville innebære at vi satte

$$(8.2) \quad y_{gj} = \begin{cases} \frac{1}{2n_g} & \text{hvis } n_{gj} = 0 \\ n_{gj}/n_g & \text{hvis } 0 < n_{gj} < n_g \\ 1 - \frac{1}{2n_g} & \text{hvis } n_{gj} = n_g. \end{cases}$$

Gart og Zweifel (1967) undersøker en del av disse metodene, men kommer til at ingen av dem kan sies å være generelt best. Haldane/Anscombe's metode kommer imidlertid relativt bra ut når det gjelder skjevhet i estimeringen av log-oddsen, unntatt når den forventede celle-frekvensen $n_g p_{g1}$ er særlig lav.

I vårt datamateriale har vi forsøkt to alternative tillegg: $\delta = 0,01$ og $\delta = 0,5$. Begrunnelsen for den første metoden er som følger: Vi ønsker å gjøre en modifikasjon som i minst mulig grad endrer observasjonsmaterialet på en vilkårlig måte, men som likevel sikrer at matrisene N og $\hat{\Sigma}$ blir inverterbare, slik at alle parametrene kan estimeres. Metoden $\delta = 0,5$ er som nevnt anbefalt i store deler av litteraturen om logit-modeller, og en sammenlikning av de to metodene for lineære såvel som for log-lineære modeller vil være av interesse.

8.4 Programutrustning

Den log-lineære modellen ble estimert ved hjelp av standardprogrammet ECTA (Everyman's Contingency Table Analyzer). Programmet gir stor valgfrihet m.h.t. hvilke samspillseffekter som skal inkluderes i modellen (første-, annen- eller evt. høyere ordens samspill). Brukeren kan dessuten fastsette tillegget δ . Programmet skriver ut både den opprinnelige og den føyede kontingenstabell, alle marginale tabeller, alle parameterestimater med estimater for standardavvikene, samt observatorene χ_p^2 og G^2 med angivelse av antall frihetsgrader. Ved hjelp av en mindre endring i programmet har vi oppnådd å få beregnet observatoren χ_N^2 (Neyman kji-kvadrat) istedenfor χ_p^2 (Pearson kji-kvadrat). Vi viser for øvrig til Clausen (1977), som inneholder en brukerbeskrivelse på norsk for programmet.

Rutiner for OLS- og 2SGLS-estimering i den lineære modell ble programmert i DATSY. Programmene skriver ut estimater for parametervektoren β , estimater for kovariansmatrisen til β -estimatoren, de føyede verdier \hat{y}_{gj} , residualene \hat{u}_{gj} , observatorene R^2 , χ_p^2 , χ_N^2 og en del andre resultater. En kan velge hvilke første-ordens samspillseffekter som skal inkluderes i modellen.

9. BEREGNINGSRESULTATER

9.1 Sammenlikning av føyingsmål

I tabellene 9.1 og 9.2 har vi samlet en del føyingsmål beregnet for et antall alternative sannsynlighetsmodeller og estimeringsmetoder anvendt på datamaterialet.

Vi har forsøkt i alt 4 ulike sannsynlighetsmodeller: (1) lineær modell uten samspillsledd, (2) lineær modell med alle 1.ordens samspillsledd mellom de uavhengige variable, (3) multinomisk logit-modell uten samspillsledd, og (4) multinomisk logit-modell med alle 1.ordens samspillsledd mellom de uavhengige variable. Dersom vi betrakter alle variable symmetrisk, kan vi gi at modellene (1) og (3) inneholder alle 3-faktor-effekter mellom de uavhengige variable, men ingen 3-faktor-effekter som inkluderer den avhengige variable. Modell (2) og (4) inneholder derimot alle mulige 3-faktor-effekter.

For de lineære modellene (1) og (2) har vi forsøkt to ulike estimeringsmetoder: OLS og 2SGLS. Modellene (3) og (4) er estimert kun ved sannsynlighetsmaksimering (ML).

For hver estimeringsmetode har vi som nevnt forsøkt med to ulike verdier på tillegget δ : 0,01 og 0,5.

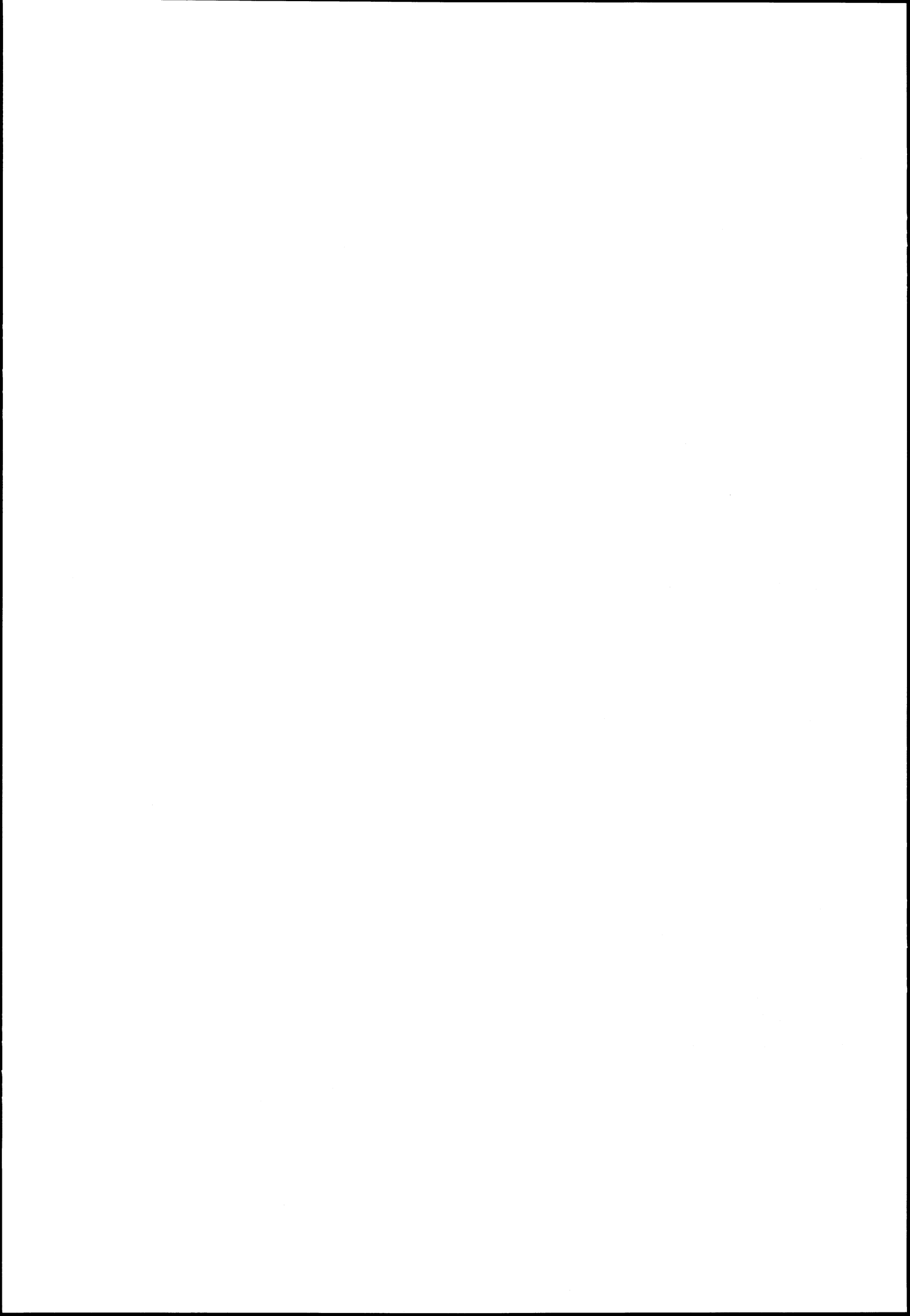
Vi har regnet ut føyingsmålene R^2 , χ_N^2 , G^2 og C , jfr. kapittel 7. Målet R^2 er regnet ut separat for de to kategoriene "deltid" og "heltid" på den avhengige variable, da det er lite naturlig å regne én determinasjonskoeffisient for hele materialet, jfr. avsnitt 7.1.1. For føyingsmålene χ_N^2 og G^2 har vi skilt ut hvor stor del av observatoren som skyldes bidrag fra celler som opprinnelig er tomme, men som bidrar fiktiv informasjon til estimeringsprosessen gjennom tillegget δ .

Merk at føyingsmålet R^2 skal maksimeres, mens målene χ_N^2 , G^2 og C skal minimeres.

Kolonnen lengst til høyre i tabell 9.1 og 9.2 gir opplysning om hvor mange føyede andeler \hat{y}_{gj} (av i alt 135) som faller utenfor intervallet $[0,1]$.

La oss først betrakte tilfellet $\delta = 0,01$ /ingen samspillsledd (tabell 9.1). OLS/lineær modell skiller seg lite fra ML/logit-modell i henhold til føyingsmålene R^2 og χ_N^2 , men kommer noe dårligere ut når en ser på G^2 eller det "nøytrale" målet C . Føyingsmålet C viser at mens logit-modell gir "feilklassifisering" av 3,21 prosent av materialet, er det tilsvarende tallet for OLS/lineær modell 3,76 prosent.

Metoden 2SGLS/lineær modell gjør det imidlertid ekstremt dårlig i dette tilfellet. Det er lett å innse hvorfor. Metoden går ut på minimering av χ_N^2 , alle celler medregnet, og på en slik måte at celler med ekstremt lave observerte relative hyppigheter y_{gj} får meget høy vekt - dette framgår klart av formelen (4.91). Mens OLS eller ML/logit tillater de 23 tomme cellene å bidra nærmere 90 prosent til føyingsmålet χ_N^2 , tvinger 2SGLS denne andelen med i ca. 52 prosent. De tomme cellene "styrer" faktisk hele estimeringsprosedyren. Resultatet blir at føyningen i de ikke-tomme cellene blir dårlig: bidraget til χ_N^2 fra disse cellene er rundt dobbelt så høyt som ved OLS eller ML/logit; til tross for at 2SGLS nettopp går ut på minimering av χ_N^2 . Også alle de øvrige føyingsmål viser at 2SGLS i dette tilfellet er en svært svak metode.



Domencich og McFadden (1975), s. 103 går så langt som til å fraråde bruk av 2SGLS i forhold til OLS på grunn av den ekstreme vekt metoden legger på celler med lave y_{gj} . Vår konklusjon er imidlertid at 2SGLS er den beste av de to metoder såsant forutsetningene for den er oppfylt. Konklusjonen bygger på teoretiske såvel som på empiriske betraktninger.

Det mest overraskende resultat i tabell 9.1 er trolig at den lineære modellen med alle 1.-ordens samspillsledd gir langt flere føyede andeler utenfor [0,1]-intervallet enn den enklere modellen uten samspillsledd. A priori vil en vente seg færre slike tilfeller jo nærmere en kommer den mettede modell. Ved nærmere ettersyn viser det seg imidlertid at i modellen med samspillsledd, skriver samtlige slike tilfeller seg fra celler med bare 0 eller 1 observasjon. Problemet med estimerte "sannsynligheter" utenfor [0,1]-intervallet i den lineære modell synes således ikke særlig alvorlig i vårt datamateriale.

Sammenlikning av tabellene 9.1 og 9.2 antyder at vi med god samvittighet kan velge $\delta = 0.5$. Dersom en bruker metoden OLS eller ML/logit, er feilklassifikasjonsprosenten, målt ved C, ikke merkbart påvirket. Metoden 2SGLS forbedres imidlertid radikalt når δ endres fra 0,01 til 0,5.

Det gjenstår å besvare spørsmålene (i) om hvor mange samspillsledd som bør tas med i modellen i vårt konkrete tilfelle, og (ii) om hvilken av de tre teknikker (OLS, 2SGLS, ML/logit) som er å foretrekke.

Spørsmål (i) kan i prinsippet avgjøres gjennom en formell testprosedyre. Enhver hypotese om fravær av samspillsledd kan skrives på formen (4.46), mens den alternative hypotesen vil være av formen (4.47). Vi så i kapittel 4 og 5 hvordan en kan konstruere asymptotiske tester av slike hypoteser. Testobservatoren utregnes enklest som økningen i den relevante kji-kvadratobservator når vi legger på restriksjonene H_0 . I den lineære modell bruker vi Neyman kji-kvadratobservatoren (χ_N^2) og i den log-lineære modell kji-kvadratobservatoren basert på sannsynlighetsknoten (G^2).

I avsnitt 9.4 kommer vi tilbake til hvilke resultater denne testprosedyren gir anvendt på vårt materiale. I denne omgang vil vi nøye oss med et par mer upresise, intuitive resonnementer. Vi holder oss nå til tilfellet $\delta = 0,5$ og noterer at føyningen for alle 6 modeller/metoder i tabell 9.2 må sies å være bra: modellen uten samspillsledd predikerer de 90 deltid- og heltidssysselesettingsandelene ved hjelp av bare 18 uavhengige parametre, mens andelen feilklassifiserte ligger i området mellom 3 og 4 prosent. De føyede andelene avviker fra de observerte med i gjennomsnitt 2 - 2,5 prosentpoeng. Modellen med alle 1.-ordens samspillsledd har 58 uavhengige parametre, og feilklassifiserer under 1 prosent. Det kan synes som om den enkleste modellen allerede gir så bra resultat at det er liten grunn til å legge på 40 nye parametre. En modell med dummy-variable og mange samspillsledd vil dessuten raskt kunne sprengte rammen for antall variable i en del standard regnemaskinprogrammer.

Når det gjelder spørsmål (ii), er det, som en kunne vente, liten forskjell mellom de tre metoder i modellen med alle 1.-ordens samspillsledd. I modellen uten samspillsledd synes ML/logit å gi noe bedre resultat enn 2SGLS/lineær modell, som igjen er noe bedre enn OLS. Forskjellene er imidlertid ikke dramatiske. Dersom en har regnemaskinprogrammer tilgjengelig som gjør det enkelt å estimere logit-modellen ved ML, er denne å foretrekke. I mange tilfeller vil imidlertid ML-estimering være svært arbeidskrevende og kostbart i forhold til OLS/lineær modell. Det samme kan sies om metoden 2SGLS. Den relativt marginale forbedring, vurdert v.h.a. våre føyningsmål, som ML/logit eller 2SGLS representerer i forhold til OLS vil ofte ikke være umaken verd.

Vi vil imidlertid ikke avslutte denne drøftingen uten å gjøre oppmerksom på et par faremomenter ved denne bruken av føyningsmål. Valget mellom lineær eller log-lineær modell bør i prinsippet bygge på overveielser om hvilken funksjonsform som best beskriver "virkeligheten". Valget av estimeringsmetode for en gitt sannsynlighetsmodell bør bygge på informasjon om hvilken metode som estimerer parametrene med minst utvalgsfeil. Det er ikke uten videre gitt at disse kriteriene også gir minst (evt. størst) verdi på et (vilkårlig valgt) føyningsmål i et empirisk materiale.

I neste avsnitt vil vi se litt nærmere på resultatene fra den lineære modell og forsøke å vurdere de to metodene OLS og 2SGLS etter (den estimerte) presisjonen på parameterestimaterne.

I kapittel 10 vil vi gå noe nærmere inn på hvilke a priori begrunnelser som kan gis for de ulike typer sannsynlighetsmodeller (logit, probit, lineær).

9.2 Variansen til parameterestimatorene i den lineære modell. Sammenlikning mellom OLS og 2SGLS

DATSY-programmene for beregning av 2SGLS- og OLS-estimatorene skriver også ut estimater for kovariansmatrisene $\Sigma_{\hat{b}}$ og $\Sigma_{\hat{\beta}}$, jfr. (4.45) og (4.69). Estimaterne er dannet ved innsetting av de observerte andeler y_{gj} i uttrykkene (4.45) og (4.69). Diagonalelementene i disse matrisene er variansestimater for estimatorene \hat{b} og $\hat{\beta}$. Disse er gjengitt i tabellene 9.3 og 9.4 for tilfellet $\delta = 0,5$. Tabellene gjengir også et effisiensmål regnet ut som forholdet mellom de estimerte standardavvik for OLS og 2SGLS:

$$(9.1) \quad \hat{\epsilon}(b_{jk}, \hat{\beta}_{jk}) = \sqrt{\widehat{\text{var}}(\hat{\beta}_{jk}) / \widehat{\text{var}}(b_{jk})}$$

b_{jk} er bedre enn $\hat{\beta}_{jk}$ når $\hat{\epsilon}(b_{jk}, \hat{\beta}_{jk})$ er større enn 1.

Tabell 9.3. Varians- og effisiensestimater for estimatorene b_{jk} (2SGLS) og $\hat{\beta}_{jk}$ (OLS). Lineær modell uten samspillsledd, $\delta = 0,5$ *Variance and efficiency estimates for the estimators b_{jk} (2SGLS) and $\hat{\beta}_{jk}$ (OLS). Linear model without interaction terms, $\delta = 0.5$*

k	Deltid <i>Part-time</i> (j=1)			Heltid <i>Full-time</i> (j=2)		
	$10^4 \cdot \widehat{\text{var}} b_{1k}$	$10^4 \cdot \widehat{\text{var}} \hat{\beta}_{1k}$	$\hat{\epsilon}(b_{1k}, \hat{\beta}_{1k})$	$10^4 \cdot \widehat{\text{var}} b_{2k}$	$10^4 \cdot \widehat{\text{var}} \hat{\beta}_{2k}$	$\hat{\epsilon}(b_{2k}, \hat{\beta}_{2k})$
1	1.090	1.154	1.029	2.346	2.412	1.014
2	1.432	1.529	1.033	4.634	4.806	1.018
3	1.419	1.562	1.049	3.932	4.097	1.021
4	1.902	2.093	1.049	5.775	5.987	1.018
5	2.182	2.397	1.048	5.034	5.289	1.025
6	0.178	0.233	1.144	0.761	0.844	1.053
7	0.553	0.633	1.070	1.544	1.757	1.067
8	0.371	0.461	1.115	1.747	1.816	1.020
9	1.721	2.390	1.178	5.015	6.400	1.130

Tabell 9.4. Varians- og effisiensestimater for estimatorene b_{jk} (2SGLS) og $\hat{\beta}_{jk}$ (OLS). Lineær modell med alle 1.-ordens samspillsledd, $\delta = 0,5$ *Variance and efficiency estimates for the estimators b_{jk} (2SGLS) and $\hat{\beta}_{jk}$ (OLS). Linear model with all 1st order interaction terms, $\delta = 0.5$*

k	Deltid <i>Part-time</i> (j=1)			Heltid <i>Full-time</i> (j=2)		
	$10^4 \cdot \widehat{\text{var}} b_{1k}$	$10^4 \cdot \widehat{\text{var}} \hat{\beta}_{1k}$	$\hat{\epsilon}(b_{1k}, \hat{\beta}_{1k})$	$10^4 \cdot \widehat{\text{var}} b_{2k}$	$10^4 \cdot \widehat{\text{var}} \hat{\beta}_{2k}$	$\hat{\epsilon}(b_{2k}, \hat{\beta}_{2k})$
1	1.865	1.876	1.003	3.278	3.288	1.002
2	3.993	4.212	1.027	14.145	14.334	1.007
3	3.113	3.253	1.022	9.149	9.262	1.006
4	7.263	8.084	1.055	27.977	28.346	1.007
5	8.598	8.965	1.021	15.813	16.312	1.016
6	4.385	4.473	1.010	11.363	11.472	1.005
7	359.921	376.967	1.023	455.112	459.468	1.005
8	99.645	110.249	1.052	105.771	116.313	1.049
9	267.974	302.180	1.062	233.466	258.770	1.053
10	6.772	7.141	1.027	24.558	24.987	1.009
11	6.044	6.489	1.036	19.384	19.787	1.010
12	8.174	8.785	1.037	28.457	29.030	1.010
13	10.007	10.513	1.025	26.385	26.918	1.010
14	366.888	384.927	1.024	484.931	489.633	1.005
15	361.280	379.844	1.025	456.072	460.465	1.005
16	369.051	388.006	1.025	480.361	484.632	1.004
17	381.038	400.946	1.026	485.498	491.587	1.006
18	101.898	113.507	1.055	119.102	131.050	1.049
19	100.370	111.429	1.054	108.708	120.137	1.051
20	104.831	116.588	1.055	131.711	143.056	1.042

Tabell 9.4 (forts.). Varians- og effisiensestimater for estimatorene b_{jk} (2SGLS) og $\hat{\beta}_{jk}$ (OLS). Lineær modell med alle 1.-ordens samspillsledd, $\delta = 0,5$ *Variance and efficiency estimates for the estimators b_{jk} (2SGLS) and $\hat{\beta}_{jk}$ (OLS). Linear model with all 1st order interaction terms, $\delta = 0.5$*

k	Deltid(j=1)			Heltid(j=2)		
	$10^4 \cdot \widehat{\text{var}} b_{1k}$	$10^4 \cdot \widehat{\text{var}} \hat{\beta}_{1k}$	$\hat{\epsilon}(b_{1k}, \hat{\beta}_{1k})$	$10^4 \cdot \widehat{\text{var}} b_{2k}$	$10^4 \cdot \widehat{\text{var}} \hat{\beta}_{2k}$	$\hat{\epsilon}(b_{2k}, \hat{\beta}_{2k})$
21	106.874	118.168	1.052	119.561	131.182	1.047
22	375.075	412.290	1.048	429.131	464.611	1.041
23	268.119	299.837	1.057	245.483	270.262	1.049
24	280.378	313.651	1.058	287.230	312.663	1.043
25	281.506	314.800	1.057	249.668	278.348	1.056
26	1.798	2.047	1.067	9.103	9.288	1.010
27	6.510	7.146	1.098	21.703	22.342	1.015
28	8.033	10.437	1.140	28.169	30.742	1.045
29	35.064	38.704	1.051	74.611	80.606	1.039

Tabell 9.3 og 9.4 bekrefter inntrykket fra tabell 9.2: det er liten forskjell mellom OLS og 2SGLS, særlig når samspillsleddene er med. Målt med standardavviket på estimatorene er 2SGLS stort sett bare litt bedre enn OLS. Mønsteret er imidlertid helt entydig: variansestimater til 2SGLS-estimatoren er mindre enn for OLS for samtlige parametre i modellen.

Zellner og Lee (1975) finner helt tilsvarende resultater i et atskillig mindre eksperiment enn vårt.

Det er ikke bare variansen til de enkelte parameterestimatorene som er av interesse, men også variansen til lineærkombinasjoner av disse. Dersom vi f.eks. er interessert i å teste hvorvidt det er forskjell på gifte og før gifte når det gjelder sannsynligheten for å arbeide deltid, vil vi trenge et estimat for

$$(9.2) \quad \text{var}(b_{18} - b_{19}) = \text{var } b_{18} + \text{var } b_{19} - 2 \text{cov}(b_{18}, b_{19}).$$

Tabell 9.5 gir de estimerte variansene til et utvalg av slike differenseestimater, beregnet ved 2SGLS eller OLS, i modellen uten samspillsledd. Bildet er helt det samme som når en ser på enkeltparametre: 2SGLS har entydig minst varians, men forskjellen er ikke overveldende.

Tabell 9.5. Varians- og effisiensestimater for estimatorene $b_{ji} - b_{jk}$ (2SGLS) og $\hat{\beta}_{ji} - \hat{\beta}_{jk}$ (OLS), for utvalgte i, k. Lineær modell uten samspillsledd, $\delta = 0,5$ *Variance and efficiency estimates for the estimators $b_{ji} - b_{jk}$ (2SGLS) and $\hat{\beta}_{ji} - \hat{\beta}_{jk}$ (OLS), for selected i, k. Linear model without interaction terms, $\delta = 0.5$*

Variabel Variable	i	k	Deltid Part-time(j=1)			Heltid Full-time(j=2)		
			$10^4 \cdot \widehat{\text{var}}(b_{1i} - b_{1k})$	$10^4 \cdot \widehat{\text{var}}(\hat{\beta}_{1i} - \hat{\beta}_{1k})$	$\hat{\epsilon}(b_{1i} - b_{1k}, \hat{\beta}_{1i} - \hat{\beta}_{1k})$	$10^4 \cdot \widehat{\text{var}}(b_{2i} - b_{2k})$	$10^4 \cdot \widehat{\text{var}}(\hat{\beta}_{2i} - \hat{\beta}_{2k})$	$\hat{\epsilon}(b_{2i} - b_{2k}, \hat{\beta}_{2i} - \hat{\beta}_{2k})$
A	2	3	0.542	0.674	1.115	3.172	3.289	1.018
	2	4	1.054	1.230	1.080	4.513	5.308	1.085
	2	5	1.357	1.587	1.081	4.555	4.763	1.023
	3	4	0.625	0.649	1.019	2.412	2.467	1.011
	3	5	0.932	1.012	1.042	1.907	1.998	1.024
	4	5	1.425	1.514	1.031	3.731	3.875	1.019
U	6	7	0.504	0.569	1.063	1.343	1.554	1.076
S	8	9	1.444	2.014	1.181	3.799	5.103	1.159

Vi kan av dette likevel ikke slutte kategorisk at 2SGLS er "bedre" enn OLS, fordi bare OLS er forventningsrett. For å komme fra variansen til bruttovariansen må en legge til kvadratet av skjevheten, jfr. formel (A.52). Vårt eksperiment kan ikke gi svar på hvor stor skjevhet som gjelder for 2SGLS. Det er derfor i prinsippet tenkelig at OLS i små utvalg gir lavere bruttovarians enn 2SGLS.

9.3 Hva tjener vi på å bruke en sannsynlighetsmodell?

Et av utgangspunktene for dette notatet var vår påstand - i kapittel 2 - om at vi, ved å bruke en sannsynlighetsmodell og estimere alle celler i tabellen simultant, kunne gjøre det bedre enn separate cellegjennomsnitt. Dette vil gjelde så sant sannsynlighetsmodellen er (ialfall tilnærmet) "riktig" valgt, og så sant den "sanne" bakenforliggende struktur er enklere enn den mettede modell.

Beregningene gir et visst grunnlag for å undersøke denne påstanden empirisk.

Variansen til de separate cellegjennomsnitt y_{gj} er gitt ved (4.14). En konsistent estimator av denne variansen er

$$(9.3) \quad \widehat{\text{var}}(y_{gj}) = \frac{1}{n_g} y_{gj}(1-y_{gj}).$$

Variansen til 2SGLS-estimatene \hat{y}_{gj} ($g=1,2,\dots,s$; $j=1,2,\dots,r-1$) er gitt ved

$$(9.4) \quad \hat{\Sigma}_{\hat{y}} = \text{Cov}(\hat{y}) = \text{Cov}(Z\hat{b}) = Z[\text{Cov}(\hat{b})]Z'.$$

En konsistent estimator av denne kovariansmatrisen er

$$(9.4) \quad \hat{\Sigma}_{\hat{y}} = Z\hat{\Sigma}_{\hat{b}}Z' = Z(Z'\hat{\Sigma}^{-1}Z)^{-1}Z'.$$

Diagonalelementene i $\hat{\Sigma}_{\hat{y}}$ er konsistente estimatører av $\text{var}(\hat{y}_{gj})$ ($g=1,2,\dots,s$; $j=1,2,\dots,r-1$).

Variansestimater for utfall r (basisutfallet) beregnes slik:

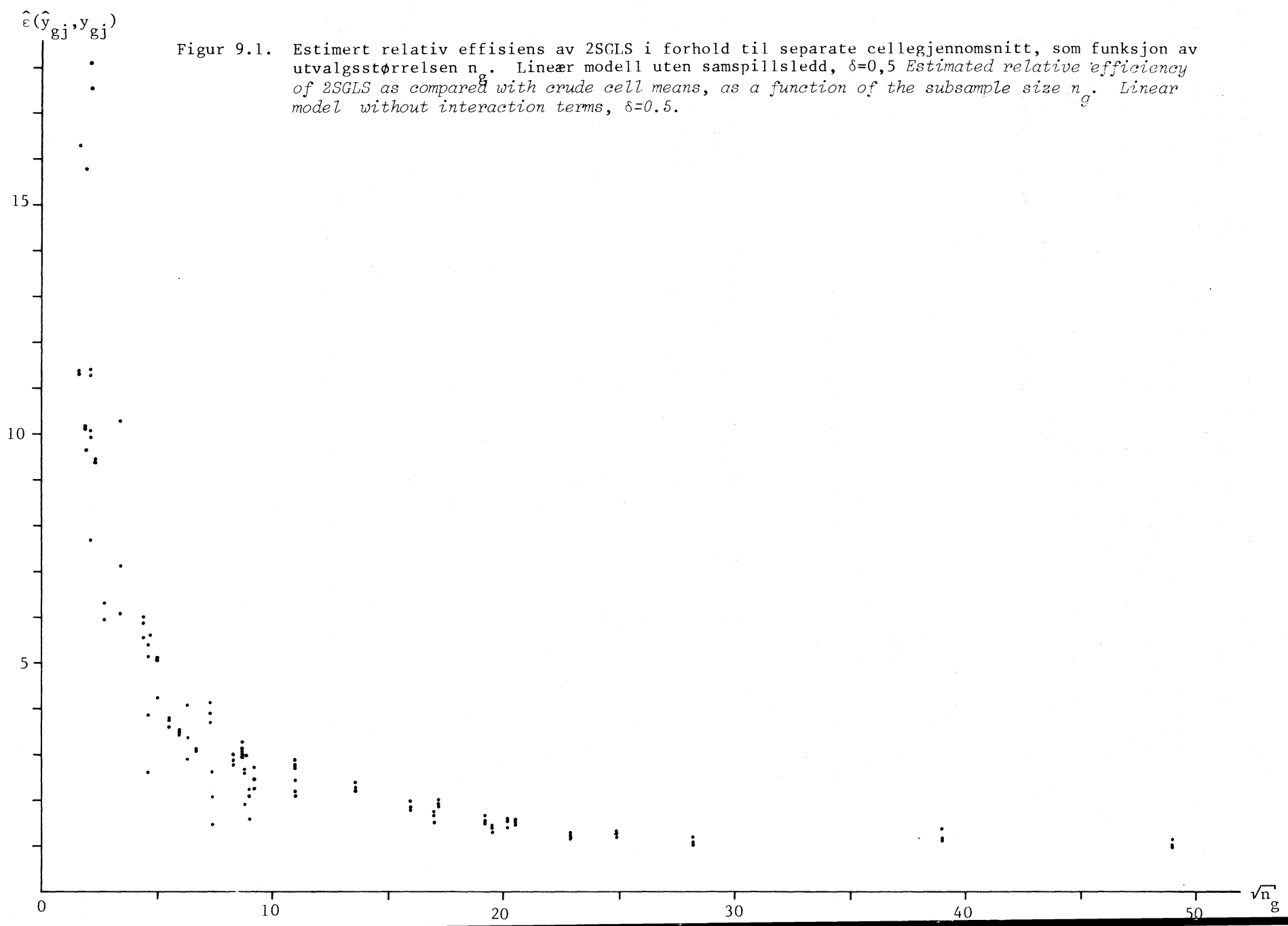
$$(9.5) \quad \hat{y}_{\cdot r} = \hat{y} - \sum_{j=1}^{r-1} \hat{y}_{\cdot j} = \hat{y} - [i_{r-1}' \otimes I_s] \hat{y}$$

dvs.

$$(9.6) \quad \begin{aligned} \widehat{\text{Cov}}(\hat{y}_{\cdot r}) &= [i_{r-1}' \otimes I_s] \hat{\Sigma}_{\hat{y}} [i_{r-1} \otimes I_s] \\ &= [i_{r-1}' \otimes I_s] Z(Z'\hat{\Sigma}^{-1}Z)^{-1}Z' [i_{r-1} \otimes I_s]. \end{aligned}$$

Matrisen $\hat{\Sigma}_{\hat{y}}$ består av $(r-1)^2$ blokker av dimensjon $(s \times s)$. Ifølge likning (9.6) får vi en konsistent estimator av kovariansmatrisen til $\hat{y}_{\cdot r}$ ved rett og slett å addere sammen disse blokkene.

Figur 9.1 gjengir noen hovedresultater av disse beregningene, dvs. vi nøyer oss med å sammenlikne variansestimater og ser her bort fra kovariansene. Hver (ikke-tomme) celle (g,j) i datamaterialet er representert ved ett punkt i figuren. Langs x-aksen har vi målt $\sqrt{n_g}$, som er omvendt proporsjonal med standardavviket til y_{gj} . Langs y-aksen måler vi den relative effisiens $\hat{\epsilon}(\hat{y}_{gj}, y_{gj})$ av y_{gj} i forhold til \hat{y}_{gj} , beregnet som forholdet mellom de estimerte standardavvik for y_{gj} og \hat{y}_{gj} . y_{gj} vil alltid være en konsistent estimator for den sanne sannsynlighet p_{gj} . \hat{y}_{gj} er konsistent for p_{gj} forutsatt at vi har brukt "riktig" sannsynlighetsmodell i estimeringen, både når det gjelder funksjonsform (lineær/log-lineær) og forekomst av samspillsledd.



Figur 9.1. Estimert relativ effisiens av 2SGLS i forhold til separate cellegjennomsnitt, som funksjon av utvalgsstørrelsen n_g . Lineær modell uten samspillsledd, $\delta=0,5$ Estimated relative efficiency of 2SGLS as compared with crude cell means, as a function of the subsample size n_g . Linear model without interaction terms, $\delta=0.5$.

Som vi ser av figuren er den potensielle gevinsten ved bruk av sannsynlighetsmodell mindre jo flere observasjoner vi har i gruppe g. Dette er et resultat vi måtte vente. Dersom gruppe g har en stor andel av det samlede observasjonsmaterialet, får vi en relativt liten økning i informasjonmengden ved å trekke inn hele resten av materialet i estimeringen. For grupper med svært få observasjoner blir imidlertid den relative økning i informasjonmengden meget stor.

Figur 9.1 gjelder dersom den sanne sannsynlighetsmodellen er lineær uten samspillsledd. Dersom den sanne modellen er lineær med alle 1.-ordens samspillsledd, blir presisjonsgevinsten betydelig mindre når vi bruker (den sanne) modellen. Dersom vi bruker en annen modell enn den sanne, er modellestimatene ikke engang konsistente. Bruk av sannsynlighetsmodell er således en avveining mellom den potensielle presisjonsgevinst vi får ved å velge rett modell, og det nokså uberegnelige tapet vi risikerer ved å velge gal modell.

9.4 Hypoteseprøving

9.4.1 Log-lineær modell

Vi har regnet ut testobservatorene $G^2(H_f)$ og $G^2(H_f|H_g)$ (jfr. avsnitt 5.5) for en rekke ulike hierarkiske hypoteser H_f og H_g , anvendt på det empiriske materialet. Resultatene er oppsummert i tabellene 9.6 - 9.8.

Tabell 9.6. Testresultater for den log-lineære modell. Kji-kvadratobservatoren (G^2), antall frihetsgrader (v_f) og signifikanssannsynlighet ($p(f)$), ved betinget og ubetinget testing. $\delta = 0$
Testing results for the log-linear model. The likelihood ratio chi-square statistic (G^2), degrees of freedom (v_f), and p-value ($p(f)$), for conditional and unconditional testing.
 $\delta = 0$

Hypotese (H_f) Hypothesis	$G^2(H_f)$	v_f	$p(f)$	H_g	$G^2(H_f) - G^2(H_g)$	$v_f - v_g$	$p(f g)$
$H_1 : \lambda^{TSUA} = 0$	36.61	32	0.263				
$H_2 : \lambda^{TSUA} = \lambda^{TSU} = 0$	53.33	40	0.077	H_1	16.72	8	0.033
$H_3 : \lambda^{TSUA} = \lambda^{TSA} = 0$	67.40	48	0.034	H_1	30.79	16	0.014
$H_4 : \lambda^{TSUA} = \lambda^{TUA} = 0$	93.15	48	0.000	H_1	56.54	16	0.000
$H_5 : \lambda^{TSUA} = \lambda^{TSU} = \lambda^{TSA} = 0$	90.63	56	0.003	H_1	54.02	24	0.000
				H_2	37.30	16	0.002
				H_3	23.23	8	0.003

Tabell 9.7. Testresultater for den log-lineære modell. Kji-kvadratobservatoren (G^2), antall frihetsgrader (ν_f) og signifikanssannsynlighet ($p(f)$), ved betinget og ubetinget testing.

$\delta = 0,01$ Testing results for the log-linear model. The likelihood ratio chi-square statistic (G^2), degrees of freedom (ν_f), and p-value ($p(f)$), for conditional and unconditional testing. $\delta = 0.01$

Hypotese (H_f)	Hypothesis	$G^2(H_f)$	ν_f	$p(f)$	H_g	$G^2(H_f) - G^2(H_g)$	$\nu_f - \nu_g$	$p(f g)$
H_1	$\lambda^{TSUA} = 0$	35.82	32	0.294				
H_2	$\lambda^{TSUA} = \lambda^{TSU} = 0$	52.43	40	0.090	H_1	16.61	8	0.034
H_3	$\lambda^{TSUA} = \lambda^{TSA} = 0$	66.47	48	0.040	H_1	30.65	16	0.015
H_4	$\lambda^{TSUA} = \lambda^{TUA} = 0$	91.96	48	0.000	H_1	56.14	16	0.000
H_5	$\lambda^{TSUA} = \lambda^{TSU} = \lambda^{TSA} = 0$	89.59	56	0.003	H_1	53.77	24	0.000
					H_2	37.16	16	0.002
					H_3	23.12	8	0.003

Tabell 9.8. Testresultater for den log-lineære modell. Kji-kvadratobservatoren (G^2), antall frihetsgrader (ν_f) og signifikanssannsynlighet ($p(f)$), ved betinget og ubetinget testing.

$\delta = 0.5$ Testing results for the log-linear model. The likelihood ratio chi-square statistic (G^2), degrees of freedom (ν_f), and p-value ($p(f)$), for conditional and unconditional testing. $\delta = 0.01$

Hypotese (H_f)	Hypothesis	$G^2(H_f)$	ν_f	$p(f)$	H_g	$G^2(H_f) - G^2(H_g)$	$\nu_f - \nu_g$	$p(f g)$
H_1	$\lambda^{TSUA} = 0$	27.46	32	0.696				
H_2	$\lambda^{TSUA} = \lambda^{TSU} = 0$	41.10	40	0.422	H_1	13.46	8	0.092
H_3	$\lambda^{TSUA} = \lambda^{TSA} = 0$	60.20	48	0.111	H_1	32.74	16	0.008
H_4	$\lambda^{TSUA} = \lambda^{TUA} = 0$	80.88	48	0.002	H_1	53.42	16	0.000
H_5	$\lambda^{TSUA} = \lambda^{TSU} = \lambda^{TSA} = 0$	80.66	56	0.017	H_1	53.20	24	0.001
					H_2	39.56	16	0.001
					H_3	20.46	8	0.009

Tabell 9.6 gir resultatene for tilfellet $\delta = 0$ (intet tillegg i tomme celler). Hypotesen H_1 sier at alle 4-faktor-effekter er lik null. (Toppskriften TSUA står for arbeidstid (T), sivilstand (S), utdanning (U) og alder (A).) I den assymmetrisk formulering tilsvarer dette ingen 2.-ordens samspill mellom S, U og A i hvordan disse påvirker T. Dette er den øverste hypotesen i "hierarkiet" - bare den mettede modell er mindre streng. Vi får her en signifikanssannsynlighet (p-verdi) på 0,263, dvs. at H_1 ikke kan forkastes på noe nivå lavere enn 26,3 prosent.

Oppmuntret av dette resultatet vandrer vi ett trinn ned i hierarkiet, og prøver hypotesen H_2 : ingen 3-faktor-effekter mellom T, S og U. Testet mot den mettede modell blir p-verdine her 7.7 prosent. Siden vi imidlertid ikke har kunnet forkaste H_1 er det rimelig å regne denne som sann. Ved å betinge m.h.p. H_1 (teste mot H_1) blir p-verdien 3,3 prosent, og vi kan forkaste H_2 på 5 prosent nivå, men ikke på 1 prosent nivå.

La oss for moro skyld bestemme oss for 1 prosent nivå og teste enda noen flere hypoteser. Hypotesene H_3 og H_4 er "sideordnet" H_2 i hierarkiet, idet alle tre innebærer at ett sett av 3-faktor-effekter er lik null. Heller ikke H_3 blir forkastet på 1 prosent nivå, verken ved betinget eller ubetinget testing. H_4 blir imidlertid forkastet.

Testet hver for seg blir altså verken H_2 eller H_3 forkastet. Kan nå begge hypoteser aksepteres simultant? Her blir svaret nei. Hypotesen $H_5 (=H_2 \cap H_3)$ gir p-verdi på 0.3 prosent testet mot den mettede modell, og mindre enn 1 promille ved betinget testing mot H_1 .

Her står vi ved et lite paradoks. Verken H_2 eller H_3 blir forkastet hver for seg, men materialet gir likevel ikke grunnlag for å påstå at begge hypoteser er sanne. Siden vi altså bare kan påstå en av hypotesene samtidig, er det naturlig å velge den som gav høyest p-verdi, nemlig H_2 .

Vi aksepterer altså H_1 og H_2 , men forkaster H_3 , H_4 og H_5 på 1 prosents nivå.

La oss for ordens skyld presisere at testene bare er gyldige dersom (1) kji-kvadrattilnærmingen er holdbar, og (2) antall frihetsgrader er riktig kalkulert. Strengt tatt er antakelig ingen av disse forutsetningene oppfylt her, idet det forventede antall tilfeller ($\hat{n}_{g,j}$) i svært mange celler er mindre enn 1, og vi dessuten burde ha korrigert antall frihetsgrader for forekomsten av tomme celler. Hensikten med tabell 9.6 har imidlertid vært å illustrere framgangsmåten ved trinnvis betinget testing av hierarkiske hypoteser.

Tabellene 9.7 og 9.8 gir testresultater for tilfellene $\delta = 0.01$ og $\delta = 0.5$. Tabellene illustrerer at slike vilkårlige tillegg kan ha avgjørende betydning for utfallet av testene. Siden tillegget ikke er nødvendig for å regne ut G^2 , er det vel mest naturlig å sette det så lavt som mulig, helst lik null.

9.4.2 Lineær modell

Mens vi i den log-lineære modell prøver en hypotese H_f ved hjelp av observatoren $G^2(H_f)$, bruker vi i den lineære modell Neyman kji-kvadrat-observatoren $\chi_N^2(H_f)$. Vi konstaterte dessuten (setning 4.2) at vi også i den lineære modell kan konstruere tester betinget med hensyn på en mindre restriktiv hypotese H_g , ved å bruke differansen $\chi_N^2(H_f) - \chi_N^2(H_g) (= \varepsilon(b_0) - \varepsilon(b))$ (jfr. (4.99)). Denne differansen er asymptotisk kji-kvadrat-fordelt under H_f med $v_f - v_g$ frihetsgrader.

Hver av hypotesene H_1, \dots, H_5 fra tabell 9.6 kan oversettes til tilsvarende hypoteser om parametervektoren β i den assymetriske lineære modell, selv om tolkningen av hver hypotese ikke blir akkurat den samme. "Oversettelsene" framgår av forspalten i tabell 9.9. Det er underforstått at ingen av hypotesene tillater 2.-ordens samspillsledd. Hypotesen H_1 betegner som før en modell med alle 1.-ordens samspill, men ingen 2.-ordens samspill i virkningen på den avhengige variable T . H_2 er som H_1 , bortsett fra at samspillet $S \times U$ forutsettes lik null, osv., jfr. tabell 9.6.

Tabell 9.9. Testresultater for den lineære modell. Kji-kvadratobservatoren (χ_N^2), antall frihetsgrader (v_f) og signifikanssannsynlighet ($p(f)$), ved betinget og ubetinget testing. $\delta = 0,5$

Testing results for the linear model. The Neyman chi-square statistic (χ_N^2), degrees of freedom (v_f), and p-value ($p(f)$), for conditional and unconditional testing. $\delta = 0.5$

Hypotese (H_f)	Hypothesis	$\chi_N^2(H_f)$	v_f	$p(f)$	H_g	$\chi_N^2(H_f) - \chi_N^2(H_g)$	$v_f - v_g$	$p(f g)$
H_1	alle 1.ordens samspill <i>all 1st order interaction terms</i>	26.25	32	0.752				
H_2	ingen samspill $S \times U$ <i>no interaction $S \times U$</i>	42.68	40	0.357	H_1	16.43	8	0.037
H_3	ingen samspill $S \times A$	66.90	48	0.037	H_1	40.65	16	0.001
H_4	ingen samspill $U \times A$	87.09	48	0.000	H_1	60.84	16	0.000
H_5	ingen samspill $S \times U, S \times A$	87.14	56	0.005	H_1	60.89	24	0.000
					H_2	44.46	16	0.000
					H_3	20.24	8	0.009

I den symmetriske log-lineære modell kan f.eks. H_2 tolkes som en bestemt form for betinget statistisk uavhengighet mellom variablene T, S og U. I den assymmetriske lineære modell er tolkningen en annen. Her betyr H_2 at variablene S og U virker additivt på T, og slik at endringen i T som følge av en bestemt endring i S, er konstant og like stor uansett verdien av U. S og U er uavhengige ikke i noen sannsynlighetsteoretisk forstand, men slik at virkningen av S er funksjonelt uavhengig av U, og omvendt. Vi står her faktisk overfor to måter å bruke begrepet "uavhengig" på, som har lite med hverandre å gjøre.

Tabell 9.9 gir forøvrig testresultater for den lineære modell tilsvarende dem vi beregnet for den log-lineære modell. På 1 prosent nivå blir H_2 (intet samspill mellom sivilstand og utdanning) ikke forkastet. H_3 , derimot, forkastes på dette nivået straks vi betinger m.h.p. H_1 . Også H_4 forkastes.

9.4.3 Sammenlikning mellom lineær og log-lineær modell

I tabell 9.10 har vi stilt sammen testresultatene for tilsvarende hypoteser i den lineære og den log-lineære modellen. For den log-lineære modellen har vi ført opp resultatene for begge tilfellene $\delta = 0$ og $\delta = 0,5$.

Stort sett ser det ut til at tilsvarende hypoteser gir noenlunde samme testresultat i begge modelltyper. Velger vi 1 prosent nivå gir de betingede testene ulikt resultat i bare ett tilfelle: H_3 (betinget m.h.p. H_1) forkastes i den lineære modell, og også i den log-lineære modell når $\delta = 0,5$, men ikke når $\delta = 0$. Velger vi 5 prosent nivå, er det fortsatt bare ett tilfelle av manglende samsvar: H_2 (betinget m.h.p. H_1). Her er det imidlertid for $\delta = 0$ at den log-lineære modellen gir samme resultat som den lineære. I denne sammenheng ser det altså ut til at verdien på tillegget δ har vel så stor betydning som valget av sannsynlighetsmodell.

Tabell 9.10. Sammenlikning av signifikanssannsynligheter for tilsvarende tester i den lineære og den log-lineære modell *Comparison of p-values for analogous tests in the linear and log-linear models*

Hypotese <i>Hypothesis</i> H_f	p(f) - Ubetinget testing <i>Unconditional testing</i>			Hypotese <i>Hypothesis</i> H_g	p(f g) - Betinget testing <i>Conditional testing</i>		
	Log-lineær modell		Lineær modell		Log-lineær modell		Lineær modell
	$\delta=0$	$\delta=0.5$	$\delta=0.5$		$\delta=0$	$\delta=0.5$	$\delta=0.5$
H_1	0.263	0.696	0.752				
H_2	0.077	0.422	0.357	H_1	0.033	0.092	0.037
H_3	0.034	0.111	0.037	H_1	0.014	0.008	0.001
H_4	0.000	0.002	0.000	H_1	0.000	0.000	0.000
H_5	0.003	0.017	0.005	H_1	0.000	0.001	0.000
				H_2	0.002	0.001	0.000
				H_3	0.003	0.009	0.009

1) Hypotesene er definert i tabell 9.6

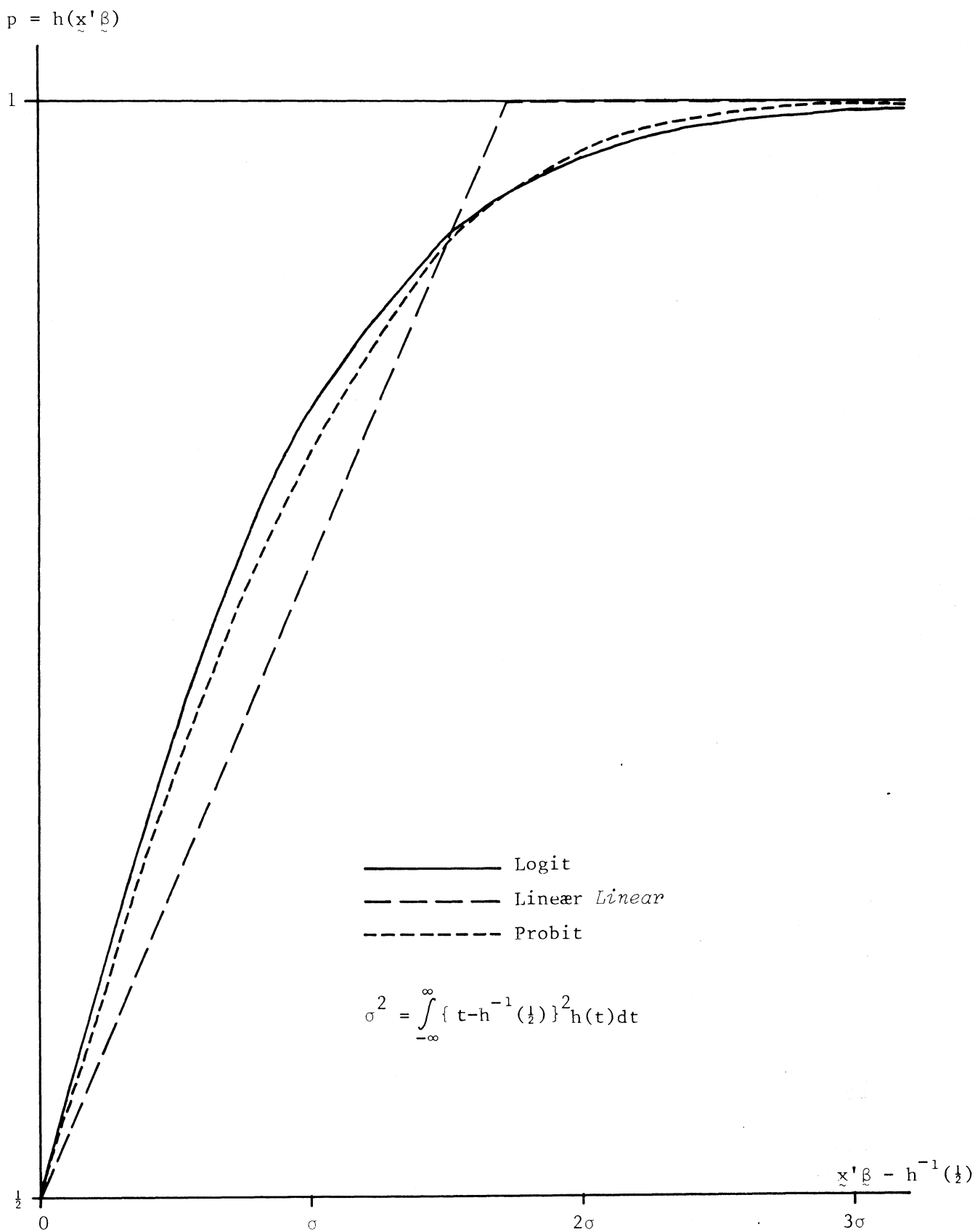
1) *The hypotheses are specified in table 9.6.*

10. FORSØK PÅ EN SYNTSE

10.1 A priori-argumenter

I figur 10.1 har vi tegnet inn de tre sannsynlighetsfunksjonene lineær, logit og probit, henholdsvis (4.5), (5.1) og (6.1). Alle kurvene er symmetriske rundt $p = 0,5$, så vi har nøyet oss med å tegne inn øvre halvdel av $[0,1]$ - intervallet. Måleenheten på x-aksen er satt lik standardavviket (σ) tilsvarende de respektive kumulative fordelingsfunksjoner: $1/(2\sqrt{3})$ for den lineære funksjon (4.5),

Figur 10.1. Forløpet av den lineære, logit og probit sannsynlighetsfunksjonene i intervallet fra $\frac{1}{2}$ og oppover *The linear, logit, and probit quantal response curves, for $p=0.5$ and above*



$\pi/\sqrt{3}$ for logit-funksjonen (5.1) og 1 for probit-funksjonen (6.1). Nullpunktet på x-aksen er den verdi ($h^{-1}(\frac{1}{2})$) der sannsynligheten er lik en halv, dvs. $\frac{1}{2}$ for den lineære modell og 0 for logit- og probit-modellen. Kurvene har derfor i en viss forstand samme skala og er direkte sammenliknbare.

Dersom den lineære kurven skal oppfattes som en sannsynlighetsmodell, må vi vedta at sannsynligheten er 1 når $\underline{x}'\underline{\beta} > 1$, jfr. (4.5). Sannsynlighetskurven har derfor knekkpunkter for $p = 0$ og $p = 1$.

Kan en si noe a priori om hvilken av de tre funksjonsformer som er "riktigst"? I spesielle tilfelle kan det tenkes at teorien foreskriver en bestemt form på funksjonen $h(\underline{x}'\underline{\beta})$. I disse tilfellene bør en selvsagt bruke nettopp denne funksjonsformen. Som regel vil en imidlertid ikke ha sterke teoretiske grunner til å velge en bestemt funksjonsform, og valget må baseres på mer intuitive rimelighetsbetraktninger.

Som et bilde på "virkeligheten" synes den lineære sannsynlighetsfunksjon med knekkpunkter svært lite rimelig. I nesten alle empiriske datamaterialer vil en kunne observere at virkningen på p av en liten økning $\Delta(\underline{x}'\underline{\beta})$ i argumentet ikke brått blir lik null ved en bestemt verdi på $\underline{x}'\underline{\beta}$, men avtar gradvis når sannsynligheten nærmer seg 1. Ofte er det heller ikke rimelig å anta at $p = 0$ eller $p = 1$ for noen verdi av argumentet.

En kan innvende at den lineære funksjon må betraktes som en approksimasjon til den "riktige" relasjonen. I området mellom $p = 0,1$ og $p = 0,9$ har både logit- og probit-funksjonen svært liten krumning, og den rette linjen burde kunne gi en bra tilnærming. Dette er viktig dersom det ikke forekommer observasjoner utenfor intervallet $[0,1, 0,9]$. Dersom det likevel forekommer slike observasjoner, vil disse i alminnelighet tvinge det lineære kurveestimatet til et mindre bratt forløp, slik at linjen heller ikke i området $[0,1, 0,9]$ vil gi god tilnærming til f.eks. logit-funksjonen, spesielt vil såkalte "slengere" ("outliers") kunne ha ødeleggende effekt på estimatene i den lineære modell.

Med andre ord, gitt at "virkeligheten" er karakterisert ved en gradvis utflatende kurve når sannsynlighetene nærmer seg 0 eller 1, vil en lineær tilnærming som regel gi dårlig tilpasning.

Den lineære funksjon vil likevel oppfylle alle konsistens- og rimelighetskrav dersom mulighetsområdet for argumentet $\underline{x}'\underline{\beta}$ er begrenset til intervallet mellom 0 og 1. Slik vil det f.eks. være i (det nokså uinteressante) tilfellet med mettet modell og bare kvalitative uavhengige variable.

Kan vi si noe a priori om valget mellom logit og probit? Berkson (1951) gikk på et tidlig stadium inn for logit-modellen. Han fant en rekke fysiske og kjemiske prosesser der den logistiske sannsynlighetsfordelingen kunne begrunnes som den a priori riktige. I samfunnsvitenskapen støter en sjelden på slike situasjoner. En kan imidlertid merke seg at probit-funksjonen stiger noe svakere i området rundt $p = \frac{1}{2}$, men til gjengjeld går raskere mot asymptotene $p = 0$ og $p = 1$. Dette er en refleks av det velkjente faktum at halene i normalfordelingen er svært tynne (tettheten går ekstremt raskt mot null når vi fjerner oss fra forventningsverdien). De fleste empiriske fordelinger en møter i praksis har antakelig tykkere haler enn normalfordelingen. Dette kan være et argument for å foretrekke logit framfor probit. Det må likevel presiseres at forskjellen mellom logit og probit-funksjonene i praksis er meget liten. I alminnelighet vil det nok ikke være avgjørende a priori grunner for å velge verken den ene eller den andre, slik at vi med relativt god samvittighet kan bruke den funksjonsform som gir det letteste regnearbeidet.

Et moment som kan komme inn i valget mellom de ulike modeller, er hvilken tolkning som kan tillegges de ulike parametre og hypoteser om disse parametrene. Vi har sett at i den log-lineære modell kan parametrene uttrykkes som funksjoner av bestemte "log-odds", og hierarkiske hypoteser om parametre kan tolkes som en form for statistisk uavhengighet. En slik egenskap har den lineære modellen ikke. Til gjengjeld er tolkningen av de enkelte parametre her ekstremt enkel og direkte: hver parameter (β_{jk}) uttrykker økningen i sannsynlighet (p_{gj}) for et bestemt utfall (j) når en gitt uavhengig variabel (x_{gk}) øker med én enhet. Tolkningen blir særlig lettfattelig nettopp fordi sammenhengen er lineær og additiv: énenhets økning i x_{gk} betinger, på godt og ondt, samme økning β_{jk} i p_{gj} uansett hvilket nivå x_{gk} (eller p_{gj}) har i utgangspunktet. Det er grunn til å tro at denne enkelhet i tolkningen av parametrene er en viktig grunn til at den lineære modellen fortsatt nyter stor popularitet.

Figur 10.2. Oversikt over noen mulige estimeringsmetoder for ulike sannsynlighetsmodeller med kvalitativ avhengig variabel, med henvisning til avsnitt i notatet.¹⁾ *Survey of feasible methods of estimation for various qualitative response models, with reference to the relevant sections of this paper.*¹⁾

Sannsynlighetsmodell <i>Probability model</i>	Type uavhengige variable <i>Type of independent variables</i>	Estimeringsmetode <i>Method of estimation</i>			
		OLS	2SGLS	IGLS	ML
Lineær <i>Linear</i>	Kvalitative <i>Qualitative</i>	4.5 (=max R^2) 2.3	4.3, 4.7 (=min x_N^2) 2.2	4.3 2.2	4.6 2.4
	Kvantitative <i>Quantitative</i>	4.5 (=max R^2) 2.3		4.3 2.2	4.6 2.4
Log-lineær <i>Log-linear (logit)</i>	Kvalitative		5.3, 5.6 (=min x_L^2) 3.2	5.3 3.2	5.4 (=min G^2) 3.3
	Kvantitative				5.4.1 3.3
Probit	Kvalitative		6 (=min x_{NO}^2)		6
	Kvantitative				6

Tegnforklaring:
Legend:



Gjennomførbare metoder
Feasible methods



Ugjennomførbare metoder
Infeasible methods



Gjennomførbarhet avhengig av datamaterialet
Feasibility depending on the actual data

1) Numrene i kursiv viser til det engelske sammendraget.

1) *Numbers in italics refer to sections of the English summary.*

For hvert valg av sannsynlighetsmodell finnes det som regel flere metoder til estimering av parametrene. I dette notatet har vi, i større eller mindre detalj, sett på de fire metodene OLS, 2SGLS, IGLS og ML. Hver metode egner seg ikke like godt til alle sannsynlighetsmodeller eller for data på alle målenivå. I figur 10.2 har vi oppsummert hvilke metoder som kan brukes ved ulike kombinasjoner av modell og målenivå. Figuren gir dessuten henvisning til de(t) avsnitt i rapporten der metoden er forklart, og i noen tilfeller til det relevante føyningsmål.

Dersom sannsynlighetsmodellen er ikke-lineær og forklaringsfaktorene kvantitative, slik at datamaterialet består av (ugrupperte) enkeltobservasjoner, er sannsynlighetsmaksimering (ML) den eneste av de fire metoder som kan anvendes. Teknikken IGLS/lineær modell kan anvendes når datasettet er eller blir manipulert slik at ingen grupper eller individer får predikert sannsynligheter utenfor $[0,1]$ -intervallet.

Figuren gir ikke svar på hvilken metode som har de beste fordelingsegenskaper. Matematisk-statistisk teori kan foreløpig bare gi oss de asymptotiske egenskaper. I den lineære modell er 2SGLS asymptotisk effisient og bedre enn OLS. I den multinomiske logit-modell er både 2SGLS og ML asymptotisk effisiente.

10.2 Praktiske erfaringer

Goldfeld og Quandt (1972) utfører et Monte-Carlo-eksperiment for å undersøke egenskapene til probit-modellen og den lineære sannsynlighetsmodell, når disse estimeres på ulike vis. Fem utvalg på fra 15 til 200 observasjoner genereres ved hjelp av hver av de to sanne sannsynlighetsmodeller, ved å forutsette et kjent konstantledd og to uavhengige variable med kjente parametre. For hver av de underliggende (sanne) sannsynlighetsmodeller estimeres parametrene ved hjelp av fire forskjellige metoder: (i) OLS basert på forutsetningen om lineær modell, (ii) 2SGLS/lineær modell (iii) ML/lineær modell og (iv) ML/probitmodell. Når dataene er generert av en probit-fordeling innebærer metodene (i), (ii) og (iii) åpenbart en modellspesifikasjonsfeil; det samme kan sies om metode (iv) anvendt på data generert av den lineære modell.

Når den sanne modell er lineær, gir metode (iii) ML/lineær modell minst bruttovarians. Bruttovariansen er litt større for OLS og 2SGLS, og svært meget større for ML/probit, unntatt i de minste utvalgene. Skjevheten blir imidlertid i alminnelighet noe lavere ved 2SGLS og OLS.

Når probit-modellen er den sanne fordeling, gir naturlig nok ML/probit-metoden i de fleste tilfeller minst bruttovarians. Mønsteret er imidlertid ikke helt entydig, og i noen tilfeller gir OLS eller ML/lineær modell lavere bruttovarians. Skjevheten blir likevel nokså entydig minst ved bruk av den "riktige" metoden (iv). Det mest slående resultat er imidlertid at alle metodene for estimering av probit-modellen har meget stor skjevhet. Den beste metoden (iv) gir estimater som avviker fra den sanne parameterverdien med fra 30 til 55 prosent i det minste utvalget og fra 20 til 25 prosent når utvalget vokser til 200. Metodene (i) - (iii) gir avvik på over 60 prosent. Konstantleddet overestimeres mens variabelkoeffisientene underestimeres. Fra teorien vet vi at metode (iv) er konsistent, dvs. at estimatorene konvergerer i sannsynlighet mot de sanne parameterverdier. Vi kan imidlertid konstatere (a) at ML/probit-estimatorene konvergerer bare meget langsomt, og (b) at metoder som urettmessig forutsetter lineær modell, har helt uakseptable fordelingsegenskaper.

Domencich og McFadden (1975) sammenlikner på liknende måte OLS/lineær-estimatoren med ML/logit og 2SGLS/logit, for en situasjon der den sanne sannsynlighetsmodell er en logit-funksjon med én variabel og uten konstantledd. De finner her at OLS/lineær metode underestimerer variabelkoeffisienten med over 50 prosent og at bruttovariansen er mange ganger så stor ved OLS/lineær metode som ved ML/logit eller 2SGLS/logit. OLS/lineær metode er som ventelig særlig svak når mange av de sanne sannsynlighetene nærmer seg 0 eller 1. Begge logit-estimatorene gir på den annen side gode resultater, og ML-estimatoren synes å konvergere relativt raskt. Det siste bekreftes av McFaddens (1974) studie, jfr. avsnitt 5.7.

Vi konkluderer altså at når "virkeligheten" er logit- eller probitformet, gir bruken av en lineær sannsynlighetsmodell til estimeringen meget dårlige parameterestimerer. Dette betyr imidlertid ikke nødvendigvis at de sanne sannsynlighetene $p_{gj} = h(x_{gj}' \beta_j)$ estimeres med like stor skjevhet, idet f.eks. den positive skjevhet for konstantleddet β_1 i noen grad oppveies av den negative skjevhet for β_2 .

Vårt empiriske materiale er egnet til å illustrere dette forholdet. Målt ved andelen feilklassifiserte individer i vårt materiale gir ML/logit-metoden noe bedre resultat enn 2SGLS/lineær metode, som igjen er litt bedre enn OLS/lineær metode. Dette gjelder dersom vi har lagt en relativt stiv struktur på materialet (ingen 1.-ordens samspill). Forskjellene er imidlertid heller små. I en modell med alle 1.-ordens samspill mellom de uavhengige variable er forskjellene mellom de tre metodene nærmest ubetydelige.

Her må vi likevel ta det forbehold at vårt datamateriale trolig er nokså spesielt, jfr. figur 8.1. For et datasett der en større andel av individene tilhører grupper der y_{gj} er nær 0 eller 1, vil føyningsmålene trolig gå klarere i favør av den log-lineære modellen.

10.3 Konklusjon

Valget av sannsynlighetsmodell for en kvalitativ avhengig variabel kan betraktes ut fra tre nokså forskjellige synsvinkler: (i) en puristisk holdning (ii) en pragmatisk holdning og (iii) en økonomisk holdning.

Med den puristiske holdning mener vi her det synspunkt at en sannsynlighetsmodell bør oppfylle visse krav til logisk konsistens, og at det er mulig, på teoretisk grunnlag, å utpeke én modell som den "riktige", eller iallfall å fastslå at av to modeller er én "riktigere" enn den andre. Ut i fra denne holdningen vil vi i alminnelighet forkaste den lineære sannsynlighetsmodell med knekkpunkter (4.5) som et uriktig bilde av virkeligheten. En modell som kan gi "sannsynligheter" utenfor $[0,1]$ - intervallet vil vi dessuten forkaste som logisk inkonsistent. Vi vil i alminnelighet ikke ha teoretisk grunnlag for å påstå en bestemt funksjonsform for sannsynlighetene til de(n) avhengige variable. Anvendt på mange samfunnsvitenskapelige datamaterialer framstår likevel logit- eller probit-funksjonene som mer rimelige enn den lineære modell.

Med den pragmatiske holdning mener vi her det synspunkt at det er de praktiske resultatene som teller. Samfunnsvitenskapelige data er sjelden oppstått som følge av en kjent, veldefinert stokastisk prosess. Vi bruker sannsynlighetsmodeller til analyse av dataene ikke fordi disse modellene er et riktig bilde av virkeligheten, men fordi modellene gjør det mulig å kondensere og trekke ut informasjon fra en ellers uoversiktlig datamasse. Valget av sannsynlighetsmodell blir rett og slett et spørsmål om hvilken metode som gjør dette mest effektivt. La oss f.eks. betrakte en modell som estimerer enkelte sannsynligheter til å ligge utenfor $[0,1]$ - intervallet. Denne modellen er ineffektiv bare i den forstand at den ikke utnytter informasjonen om at slike sannsynligheter ikke kan forekomme, og bare i den grad dette gjør resultatene mindre pålitelige.

Siden det, ifølge den pragmatiske holdning, ikke finnes a priori teoretiske holdepunkter for å velge mellom de ulike statistiske teknikker, må vurderingen baseres på empiri. Det pragmatiske synspunkt kan da presiseres f.eks. slik: Velg den metode som gir den beste tilpasning til dataene i vårt materiale. Denne regelen er imidlertid ikke mer presis enn at det gjenstår å fastsette et føyningsmål eller annet (vilkårlig) vurderingskriterium. Vi har sett hvordan ulike føyningsmål kan gi vidt forskjellige konklusjoner. Ofte finner vi ikke noe entydig svar.

I praksis vil derfor ofte det økonomiske synspunkt være avgjørende: Velg, under ellers like forhold, den teknikk som er minst ressurskrevende. Ressursbruken vil avhenge i det enkelte tilfelle av tilgangen til datamaskiner og i første rekke av hvilken programvare som er praktisk tilgjengelig. Vanlig lineær regresjon (OLS) vil for de fleste brukere være av de billigste alternativer. Ikke minst takket være programmet ECTA er også ML/logit en billig analysemetode, som imidlertid bare kan nyttes på grupperte data. Noe avhengig av programutrustningen vil ML på individdata ofte være svært ressurskrevende. Av de teknikker vi har forsøkt har 2SGLS vært den klart dyreste.

Dersom vi inntar en puristisk holdning er metodevalget i alminnelighet lett: Vi går inn for en multinomisk logit-modell. Denne oppfyller alle krav til logisk konsistens og er trolig den minst ressurskrevende av alle slike modeller. Både ML- og 2SGLS-estimatorene har, i motsetning til i probit-modellen, gode fordelingssegenskaper.

Med utgangspunkt i en pragmatisk holdning er svaret mindre klart. Vurdert etter vårt mest "nøytrale" føyningsmål er ML/logit den beste av de tre metoder vi har anvendt på vårt AKU-materiale. 2SGLS/lineær modell er nest best og OLS/lineær modell den dårligste. I vårt datamateriale er forskjellene imidlertid såpass små at den pragmatiske vurdering ikke frambringer særlig tungtveiende argumenter. Vurderingen vil trolig gå mer i retning av logit-modellen jo større andel av sannsynlighetene som ligger nær 0 eller 1. Kostnadsaspektet får imidlertid atskillig vekt.

1. ABSTRACT

This paper investigates and compares the properties of certain qualitative response models, notably the linear (or additive) model, the log-linear (or multinomial logit) model, and, in less detail, the binary probit model. A data set taken from the Norwegian labour force sample surveys is used to confront the linear and the log-linear models empirically.

As a by-product, we develop a method to study the properties of the so-called coefficient of determination (R^2), in a linear model with binary dependent variable. We start from the suspicion that in such situations the R^2 measure inherently tends to be low, a conjecture which is wholly confirmed by our investigations.

Equations and inequalities in this summary carry the same numbers as in the Norwegian text. Diagrams and tables are not repeated here - cf. the indexes of tables and figures on p. 11-12.

2. THE LINEAR MODEL

2.1. Notation

Let x_g be an $(m \times 1)$ vector of attributes characterizing all individuals belonging to subpopulation g , and assume we have a random sample of n_g individuals belonging to this group. Also, let n_{gj} be the number of individuals in the sample who belong to category g and who attain a value j on some qualitative response variable of interest. Throughout this paper, we refer to the first index g as indicating membership in "group" g , while the different values of j are termed "outcomes" or "alternatives". It is assumed that there are s distinct groups and r distinct alternatives, i.e. g runs from 1 to s and j from 1 to r . Finally, let $y_{gj} = n_{gj}/n_g$, i.e. y_{gj} is the proportion of respondents in group g who have "chosen" alternative j .

Note that while the response variable j is assumed to be qualitative, no such constraint is placed upon the vector of independent variables, x_g . "Qualitative" is here taken to mean "measured on the nominal or ordinal level", as opposed to "quantitative" variables, which are "measured on the interval or ratio level" (cf. Stevens (1946)).

Also, in principle no assumption is made about the size of n_g . Indeed, if x_g contains one or more quantitative variables, it is likely that no two respondents in the sample exhibit the same value on x_g , such that each n_g equals one and $s = n = \sum_{g=1}^s n_g$. Thus, in this case the sample, rather than taking the form of a contingency table $\{n_{11}, \dots, n_{sr}\}$, consists of nr binary variables y_{ij}^* ($i=1, \dots, n; j=1, \dots, r$) and n individual attribute vectors x_i^* . (Sometimes, to distinguish this situation from the replicated observations case, we write i rather than g as the first index and attach the superscript *.) Here, y_{ij}^* can take on only the two values 0 or 1.

Now, the linear model can be written as follows:

$$\begin{aligned}
 (4.12) \quad y_{gj} &= x_g^1 \beta_j + u_{gj} \\
 (4.13) \quad E(u_{gj}) &= 0 \\
 (4.14) \quad \text{var}(u_{gj}) &= \frac{1}{n_g} p_{gj}(1-p_{gj})
 \end{aligned}
 \left. \vphantom{\begin{aligned} (4.12) \\ (4.13) \\ (4.14) \end{aligned}} \right\} \begin{array}{l} (g=1,2,\dots,s; \\ j=1,2,\dots,r) \end{array}$$

$$(4.15) \quad \text{cov}(u_{gj}, u_{hk}) = \begin{cases} 0 & \text{for } g \neq h \\ -\frac{1}{n_g} p_{gj} p_{gk} & \text{for } g = h, j \neq k \end{cases}$$

where $p_{gj} = E(y_{gj}) = x'_{gj}\beta_j$. β_j is an $(m \times 1)$ vector of parameters, while u_{gj} is a random error term with zero expectation. Error terms concerning different groups g and h are assumed to be independent, while the covariance between u_{gj} and u_{gk} is, of course, negative (for $j \neq k$).

Defining the matrices

$$(4.16) \quad X = \begin{bmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_s \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{s1} & x_{s2} & \dots & x_{sm} \end{bmatrix}$$

$$(4.22) \quad Z = [I_{r-1} \otimes X] = \begin{bmatrix} X & \emptyset & \dots & \emptyset \\ \emptyset & X & \dots & \emptyset \\ \vdots & \vdots & & \vdots \\ \emptyset & \emptyset & \dots & X \end{bmatrix}$$

and the vectors

$$(4.17) \quad \underline{y}_{\cdot j} = (y_{1j} \ y_{2j} \ \dots \ y_{sj})'$$

$$(4.18) \quad \underline{u}_{\cdot j} = (u_{1j} \ u_{2j} \ \dots \ u_{sj})'$$

$$(j=1, 2, \dots, r)$$

$$(4.21) \quad \underline{y} = (\underline{y}'_{\cdot 1} \ \underline{y}'_{\cdot 2} \ \dots \ \underline{y}'_{\cdot r-1})'$$

$$(4.23) \quad \underline{\beta} = (\beta'_1 \ \beta'_2 \ \dots \ \beta'_{r-1})'$$

$$(4.24) \quad \underline{u} = (\underline{u}'_{\cdot 1} \ \underline{u}'_{\cdot 2} \ \dots \ \underline{u}'_{\cdot r-1})'$$

the model can be written more compactly as

$$(4.19) \quad \underline{y}_{\cdot j} = X \underline{\beta}_j + \underline{u}_{\cdot j} \quad (j=1, 2, \dots, r)$$

or

$$(4.20) \quad \underline{y} = Z \underline{\beta} + \underline{u}$$

Here,

$$(4.25) \quad E(\underline{u}) = \underline{0}$$

$$(4.26) \quad \Sigma = \text{Cov}(\underline{u}) = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \dots & \Sigma_{1,r-1} \\ \Sigma_{12} & \Sigma_{22} & \dots & \Sigma_{2,r-1} \\ \vdots & \vdots & & \vdots \\ \Sigma_{1,r-1} & \Sigma_{2,r-1} & \dots & \Sigma_{r-1,r-1} \end{bmatrix},$$

the blocks of Σ being defined by

$$(4.27) \quad \Sigma_{jj} = \begin{bmatrix} p_{1j}(1-p_{1j})/n_1 & 0 & \dots & 0 \\ 0 & p_{2j}(1-p_{2j})/n_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & p_{sj}(1-p_{sj})/n_s \end{bmatrix} \quad (j=1,2,\dots,r-1)$$

and

$$(4.28) \quad \Sigma_{jk} = \begin{bmatrix} -p_{1j}p_{1k}/n_1 & 0 & \dots & 0 \\ 0 & -p_{2j}p_{2k}/n_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -p_{sj}p_{sk}/n_s \end{bmatrix} \quad (j < k).$$

Here, I_k denotes the $(k \times k)$ identity matrix, \emptyset the zero matrix, and \otimes the familiar Kronecker product.

Each vector $(n_{g1}, n_{g2}, \dots, n_{gr})$ is, of course, multinomially distributed with parameters n_g and $p_{g1}, p_{g2}, \dots, p_{gr}$. Because of the obvious constraint

$$(4.29) \quad \sum_{j=1}^r y_{gj} = 1 \quad (g=1,2,\dots,s)$$

it is sufficient to work with observations on the $r-1$ first alternatives only, the last one being determined residually. We refer to this last (r 'th) outcome as the reference alternative.

One major drawback with this model is that there is no inherent guarantee that all estimated (or predicted) probabilities fall inside the unit interval. Indeed, this is one major reason why we also want to consider other possible qualitative response models, some of them rather more complicated than the linear model.

2.2. The two-stage generalized least squares (2SGLS) estimator

There are several ways of estimating the linear model. At the outset, let us note that the theoretically best linear unbiased estimator of β is given by

$$(4.40) \quad b^0 = (Z' \Sigma^{-1} Z)^{-1} Z' \Sigma^{-1} y.$$

This is the so-called GLS (generalized least squares) or Aitken estimator. It has covariance matrix given by

$$(4.41) \quad \Sigma_{b^0} = (Z' \Sigma^{-1} Z)^{-1}.$$

All other linear unbiased estimators have covariance matrices that exceed Σ_{b^0} by a positive semi-definite matrix.

The trouble with b^0 is that it cannot be computed, since Σ depends precisely on those true, unknown probabilities p_{gj} which we want to estimate. The so-called feasible Aitken estimator is defined as follows: Let $\hat{\Sigma}$ be a consistent estimate of Σ , and replace Σ by $\hat{\Sigma}$ in formula (4.40).

In this paper we consider two such feasible Aitken estimators, termed the two-stage generalized least squares (2SGLS) estimator and the iterative generalized least squares (IGLS) estimator.

The 2SGLS estimator \underline{b} is defined by

$$(4.42) \quad \underline{b} = (Z' \hat{\Sigma}^{-1} Z)^{-1} Z' \hat{\Sigma}^{-1} y,$$

where $\hat{\Sigma} = \Sigma |_{p_{gj}=y_{gj}}$. In other words, a consistent estimate of Σ is found by replacing the true, unknown probabilities p_{gj} in formula (4.27) - (4.28) by the observed relative frequencies y_{gj} . This method is, however, practicable only in the replicated observations case ($n_g > 1$). Indeed, we must require that all y_{gj} be strictly greater than zero and strictly less than one. We shall return to the question of how to handle occasional zeros in the contingency table.

The IGLS estimator is defined as follows. (1) Plug $p_{gj} = \frac{1}{r}$ ($g=1,2,\dots,s; j=1,2,\dots,r-1$) into (4.27) and (4.28) and compute \underline{b}^0 using formula (4.40). Compute the fitted values $\hat{y}^0 = Z \underline{b}^0$. \hat{y}^0 is then a consistent estimator of $(p_{11}, \dots, p_{s,r-1})'$. (2) Now analogously use the fitted values \hat{y}_{gj}^0 to form the matrix $\hat{\Sigma}^0$ and compute

$$(4.50) \quad \underline{b}^1 = (Z' (\hat{\Sigma}^0)^{-1} Z)^{-1} Z' (\hat{\Sigma}^0)^{-1} y$$

and next

$$(4.51) \quad \hat{y}^1 = Z \underline{b}^1.$$

(3) Continue by forming $\hat{\Sigma}^1$ based on \hat{y}^1 etc. until convergence. McGillivray (1970) shows that this method is consistent. Compared with 2SGLS, IGLS has the advantage of being practicable also in the non-replicated observations case, i.e. where x_g contains quantitative variables. It has, however, two major disadvantages. Firstly, we are likely to run into difficulties because not all intermediate estimates \hat{y}_{gj}^t will obey $0 < \hat{y}_{gj}^t < 1$ ($t=0,1,\dots$). Certain arbitrary ad hoc procedures must then be agreed upon to avoid negative variance estimates or possibly a (near-)singular covariance matrix estimate. Secondly, the IGLS method is likely to be considerably more expensive and computationally cumbersome than 2SGLS. For these reasons, only the 2SGLS method is studied in any greater detail here.

Following Dhrymes (1970), p. 161-166 it is easy to show that 2SGLS is asymptotically equivalent to the GLS estimator, both estimators obeying

$$(4.43) \quad \sqrt{n}(\underline{b}-\underline{\beta}) \xrightarrow{L} N(0, \lim_{n \rightarrow \infty} n(Z' \Sigma^{-1} Z)^{-1}).$$

Here, the sign \xrightarrow{L} denotes convergence in distribution (law), while in general $N(\underline{\mu}, \Sigma)$ denotes the (multivariate) normal distribution with mean $\underline{\mu}$ and covariance matrix Σ . It should be noted, however, that (4.43) is based on the assumption that as n approaches infinity, so does the size n_g of every subsample. To be specific, it is assumed that if we define

$$(4.30) \quad N = \begin{bmatrix} n_1 & 0 & \dots & 0 \\ 0 & n_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & n_s \end{bmatrix}$$

$$(4.31) \quad M = I_{r-1} \otimes N$$

and

$$(4.34) \quad P_n = \frac{1}{n} N$$

$$(4.35) \quad Q_n = I_{r-1} \otimes P_n = \frac{1}{n} M,$$

then the diagonal elements of P_n , n_g/n , converge to a set of constants as n approaches infinity:

$$(4.36) \quad \lim_{n \rightarrow \infty} P_n = P$$

$$(4.37) \quad \lim_{n \rightarrow \infty} Q_n = Q = I_{r-1} \otimes P.$$

A consistent estimate of the asymptotic covariance matrix of \underline{b} , $\hat{\Sigma}_b$, is given by

$$(4.45) \quad n\hat{\Sigma}_b = n(Z'\hat{\Sigma}^{-1}Z)^{-1} \xrightarrow{P} (Z'Q\Sigma_0^{-1}Z)^{-1}$$

where $\Sigma_0 = \Sigma \Big|_{n_g=1}$. Assume that we want to test a hypothesis of the form

$$(4.46) \quad H_0: K\underline{\beta} = \underline{\kappa}$$

against the alternative

$$(4.47) \quad H_1: K\underline{\beta} \neq \underline{\kappa}.$$

K is a $(v \times m)$ constant matrix of rank v and $\underline{\kappa}$ is a $(v \times 1)$ constant vector. Then, under H_0 , it can be shown that

$$(4.48) \quad \sqrt{n}(\underline{Kb} - \underline{\kappa}) \xrightarrow{L} N(0, \lim_{n \rightarrow \infty} nK(Z'\Sigma^{-1}Z)^{-1}K')$$

and that, furthermore,

$$(4.49) \quad (\underline{Kb} - \underline{\kappa})' [K\hat{\Sigma}_b K']^{-1} (\underline{Kb} - \underline{\kappa}) \xrightarrow{L} \chi^2(v).$$

Thus, an asymptotic level α test of H_0 vs. H_1 consists in rejecting H_0 whenever the statistic (4.49) exceeds the $(1-\alpha)$ quantile of the chi-square distribution with v degrees of freedom.

We can prove the following handy result:

Proposition 4.1. Define the Neyman (or modified Pearson) chi-square statistic

$$(4.91) \quad \chi_N^2 = \sum_{g=1}^s n_g \sum_{j=1}^r [(y_{gj} - p_{gj})^2 / y_{gj}].$$

Minimizing χ_N^2 with respect to $\underline{\beta}$ (recall $p_{gj} = x_{gj}'\beta_j$) is equivalent to 2SGLS, in the very strong sense that the two estimators are, in fact, algebraically identical.

The proof of this result is a straight-forward extension of Clickner's (1976) theorem 2.2, which deals with the special case $s = 1$.

Neyman (1949) showed that the $\min \chi_N^2$ estimator is BAN (best asymptotically normal). Consequently, this is also true of the 2SGLS estimator.

Another neat result is

Proposition 4.2. Consider again the two hypotheses H_0 and H_1 , given by (4.46) and (4.47), and the asymptotic test statistic (4.49). Let \underline{b}_0 denote the 2SGLS ($= \min \chi_N^2$) estimate of $\underline{\beta}$ computed under H_0 , and let \underline{b} be the estimator computed under H_1 . Also, denote by $\xi(\underline{b}_0)$ and $\xi(\underline{b})$ the corresponding Neyman chi-square statistics calculated under H_0 and H_1 , respectively. Then

$$(4.99) \quad \xi(\underline{b}_0) - \xi(\underline{b}) = (\underline{Kb} - \underline{\kappa})' [K(Z'\hat{\Sigma}^{-1}Z)^{-1}K']^{-1} (\underline{Kb} - \underline{\kappa}),$$

i.e. the test statistic (4.49) is equal to the difference between the two chi-square measures.

Sketch of proof: Define the Lagrange function

$$(4.100) \quad \begin{aligned} \omega(\underline{\beta}) &= \xi(\underline{\beta}) + \underline{\pi}'(K\underline{\beta} - \underline{\kappa}) \\ &= (\underline{y} - Z\underline{\beta})' \hat{\Sigma}^{-1} (\underline{y} - Z\underline{\beta}) + \underline{\pi}'(K\underline{\beta} - \underline{\kappa}), \end{aligned}$$

$\underline{\pi}$ being a vector of Lagrange multipliers. Differentiating (4.100) with respect to $\underline{\beta}$ and $\underline{\pi}$ and setting the derivatives equal to zero yield the 1st-order conditions for minimum

$$(4.101) \quad \frac{\partial \omega}{\partial \underline{\beta}} = 2Z' \hat{\Sigma}^{-1} Z \underline{\beta} - 2Z' \hat{\Sigma}^{-1} \underline{y} + K' \underline{\pi} = \underline{0}$$

$$(4.102) \quad \frac{\partial \omega}{\partial \underline{\pi}} = K\underline{\beta} - \underline{\kappa} = \underline{0}.$$

Note that \underline{b}_0 is found by solving this system as it stands, while \underline{b} is found by letting $\underline{\pi} = \underline{0}$ (no constraints) and then solving (4.101). In other words,

$$(4.103) \quad Z' \hat{\Sigma}^{-1} (\underline{y} - Z\underline{b}) = \underline{0}.$$

Now, premultiply (4.101) by $K(Z' \hat{\Sigma}^{-1} Z)^{-1}$, use $K\underline{\beta} = \underline{\kappa}$ (4.46), and obtain (under H_0)

$$(4.104) \quad \begin{aligned} \underline{\pi} &= 2 \cdot [K(Z' \hat{\Sigma}^{-1} Z)^{-1} K']^{-1} [K(Z' \hat{\Sigma}^{-1} Z)^{-1} Z' \hat{\Sigma}^{-1} \underline{y} - \underline{\kappa}] \\ &= 2 \cdot [K(Z' \hat{\Sigma}^{-1} Z)^{-1} K']^{-1} (K\underline{b} - \underline{\kappa}). \end{aligned}$$

Plugging this expression back into (4.101) and using (4.42), we find

$$(4.105) \quad \underline{b}_0 = \underline{b} - (Z' \hat{\Sigma}^{-1} Z)^{-1} K' [K(Z' \hat{\Sigma}^{-1} Z)^{-1} K']^{-1} (K\underline{b} - \underline{\kappa}).$$

Moreover,

$$(4.106) \quad \begin{aligned} \xi(\underline{b}_0) &= (\underline{y} - Z\underline{b}_0)' \hat{\Sigma}^{-1} (\underline{y} - Z\underline{b}_0) \\ &= [\underline{y} - Z\underline{b} - Z(\underline{b}_0 - \underline{b})]' \hat{\Sigma}^{-1} [\underline{y} - Z\underline{b} - Z(\underline{b}_0 - \underline{b})] \\ &= (\underline{y} - Z\underline{b})' \hat{\Sigma}^{-1} (\underline{y} - Z\underline{b}) - 2(\underline{b}_0 - \underline{b})' Z' \hat{\Sigma}^{-1} (\underline{y} - Z\underline{b}) + (\underline{b}_0 - \underline{b})' Z' \hat{\Sigma}^{-1} Z (\underline{b}_0 - \underline{b}). \end{aligned}$$

Here, the first term equals $\xi(\underline{b})$ while the second term vanishes in view of (4.103). Thus, combining (4.105) and (4.106) we derive

$$(4.107) \quad \begin{aligned} \xi(\underline{b}_0) - \xi(\underline{b}) &= (K\underline{b} - \underline{\kappa})' [K(Z' \hat{\Sigma}^{-1} Z)^{-1} K']^{-1} K(Z' \hat{\Sigma}^{-1} Z)^{-1} Z' \hat{\Sigma}^{-1} Z (Z' \hat{\Sigma}^{-1} Z)^{-1} K' \cdot [K(Z' \hat{\Sigma}^{-1} Z)^{-1} K']^{-1} (K\underline{b} - \underline{\kappa}) \\ &= (K\underline{b} - \underline{\kappa})' [K(Z' \hat{\Sigma}^{-1} Z)^{-1} K']^{-1} (K\underline{b} - \underline{\kappa}), \end{aligned}$$

and the proof is complete.

Proposition 4.2 provides us with a short-cut method of calculating the test statistic (4.49). It is often easier to calculate the two chi-square measures $\xi(\underline{b}_0)$ and $\xi(\underline{b})$ than to compute the quadratic form (4.49).

Also, it is seen that this result gives rise to a way of partitioning the Neyman chi-square statistic so as to be able to test successively a whole chain of nested hypotheses (cf. section 5.8 of this summary).

The calculation of the 2SGLS estimator \underline{b} can be considerably simplified by noting that the matrix Σ - and, by the same token, $\hat{\Sigma}$ - can be inverted analytically. To be specific,

$$(4.52) \quad \Sigma^{-1} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \dots & \Sigma_{1,r-1} \\ \Sigma_{12} & \Sigma_{22} & \dots & \Sigma_{2,r-1} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{1,r-1} & \Sigma_{2,r-1} & \dots & \Sigma_{r-1,r-1} \end{bmatrix}$$

where

$$(4.53) \quad \Sigma^{jj} = \begin{bmatrix} n_1(p_{1r}^{-1} + p_{1j}^{-1}) & 0 & \dots & 0 \\ 0 & n_2(p_{2r}^{-1} + p_{2j}^{-1}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & n_s(p_{sr}^{-1} + p_{sj}^{-1}) \end{bmatrix} \quad (j=1,2,\dots,r-1)$$

and

$$(4.54) \quad \Sigma^{jk} = \begin{bmatrix} n_1 p_{1r}^{-1} & 0 & \dots & 0 \\ 0 & n_2 p_{2r}^{-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & n_s p_{sr}^{-1} \end{bmatrix} \quad (j < k).$$

Thus, the inverse Σ^{-1} is of the same form as Σ , in the sense that both matrices consist of $(r-1)^2$ quadratic blocks of size $(s \times s)$, each block being itself a diagonal matrix. The inverse $\hat{\Sigma}^{-1}$ is of course found by substituting y_{gj} for p_{gj} in (4.53)-(4.54).

2.3. The ordinary least squares (OLS) estimator

For the sake of the argument, let us assume that the entire sample, instead of being collapsed into a contingency table, is treated as n individual observations. Each respondent is described by a set of dummy response variables $y_{i1}^*, \dots, y_{i,r-1}^*$ and a vector of independent variables x_i^* ($i=1,2,\dots,n$). Define

$$(4.58) \quad \underline{y}^* = (y_{11}^* \dots y_{n1}^* \quad y_{12}^* \dots y_{n2}^* \quad \dots \quad y_{1,r-1}^* \dots y_{n,r-1}^*)'$$

$$(4.59) \quad \underline{u}^* = (u_{11}^* \dots u_{n1}^* \quad u_{12}^* \dots u_{n2}^* \quad \dots \quad u_{1,r-1}^* \dots u_{n,r-1}^*)'$$

$$(4.60) \quad Z^* = I_{r-1} \otimes X^*$$

where

$$(4.61) \quad X^* = \begin{bmatrix} x_1^* \\ x_2^* \\ \vdots \\ x_n^* \end{bmatrix} = \begin{bmatrix} i_{n_1} \otimes x_1^* \\ \vdots \\ i_{n_s} \otimes x_s^* \end{bmatrix}$$

and \underline{i}_k denotes the $(k \times 1)$ unit vector. Note that $x_i^* = x_g$ if and only if the i 'th respondent belongs to the g 'th group.

Using these formulae, the model (4.20) can be written alternatively as

$$(4.57) \quad \underline{y}^* = Z^* \underline{\beta} + \underline{u}^*.$$

We define the ordinary least squares (OLS) estimator of $\underline{\beta}$ as

$$(4.65) \quad \hat{\underline{\beta}} = (Z^{*'} Z^*)^{-1} Z^{*'} \underline{y}^*$$

or, equivalently,

$$(4.66) \quad \hat{\underline{\beta}} = (Z' M Z)^{-1} Z' M \underline{y},$$

where M is defined by (4.31).

The OLS estimator has covariance matrix given by

$$(4.69) \quad \Sigma_{\hat{\underline{\beta}}} = (Z' M Z)^{-1} Z' M \Sigma_M Z (Z' M Z)^{-1}.$$

and asymptotic distribution given by

$$(4.71) \quad \sqrt{n} (\hat{\underline{\beta}} - \underline{\beta}) \xrightarrow{L} N(0, \lim_{n \rightarrow \infty} n \Sigma_{\hat{\underline{\beta}}}).$$

Moreover, under the hypothesis (4.46), we have, analogously to (4.49):

$$(4.73) \quad (K_{\hat{\underline{\beta}}} - \underline{\kappa})' [K_{\hat{\underline{\beta}}} K']^{-1} (K_{\hat{\underline{\beta}}} - \underline{\kappa}) \xrightarrow{L} \chi^2(\nu).$$

It is worth noting that since both M and Z are block-diagonal matrices, each subvector $\underline{\beta}_j$ of $\underline{\beta}$ might as well be estimated separately:

$$(4.78) \quad \hat{\underline{\beta}}_j = (X' N X)^{-1} X' N \underline{y}_{\cdot j} \quad (j=1, 2, \dots, r).$$

Formulae (4.65), (4.66) and (4.78) all yield algebraically identical estimates of the parameters.

Albeit each subvector $\underline{\beta}_j$ is estimated separately, it can be shown that as long as the response variables fulfill the condition

$$\sum_{j=1}^r y_{gj} \quad (\text{or } \sum_{j=1}^r y_{ij}^*) = 1 \quad (g=1, 2, \dots, s, \quad i=1, 2, \dots, n),$$

so will the corresponding fitted values $\hat{y}_{gj} = x_{gj}' \hat{\underline{\beta}}_j$ (or $\hat{y}_{ij}^* = x_{ij}^* \hat{\underline{\beta}}_j$). Thus, the estimated probabilities sum up to one for every group g and every individual i .

Dhrymes (1970), p. 159-160, shows that the OLS estimator is asymptotically inefficient compared to the 2SGLS estimator, in the sense that the asymptotic covariance matrix of $\hat{\underline{\beta}}$ exceeds that of \underline{b} by a positive definite matrix. The small sample properties of 2SGLS are not known, but it is hardly an adventurous guess that 2SGLS is superior to OLS also in smaller samples. It is, however, true that while both estimators are consistent, only OLS is unbiased. For a comparative empirical study of these estimators we refer the reader to section 5.5 of this summary.

2.4. The maximum likelihood (ML) estimator

The log-likelihood function is given by

$$(4.83) \quad \log L = \sum_{g=1}^s \left\{ \log(n_g!) + \sum_{j=1}^r [n_{gj} \log p_{gj} - \log(n_{gj}!)] \right\}$$

where

$$(4.84) \quad \left\{ \begin{array}{l} p_{gj} = \frac{x_g' \beta_j}{z_g \beta_j} \quad (j=1,2,\dots,r-1) \\ p_{gr} = 1 - \sum_{j=1}^{r-1} p_{gj} = 1 - \sum_{j=1}^{r-1} \frac{x_g' \beta_j}{z_g \beta_j} \end{array} \right\}.$$

The 1st order conditions for maximum are

$$(4.86) \quad \frac{\partial \log L}{\partial \beta_{jk}} = \sum_{g=1}^s \left[n_{gj} x_{gk} / (x_g' \beta_j) - n_{gr} x_{gk} / (1 - \sum_{j=1}^{r-1} x_g' \beta_j) \right] = 0 \quad (j=1,2,\dots,r-1; k=1,2,\dots,m),$$

where β_{jk} is the k 'th element of subvector β_j .

Note that $\log L$ depends on p_{gj} only through the logarithmic function, which is undefined for negative arguments. Hence, for every combination $(y_{g1}, \dots, y_{gr}, x_g')$ of dependent and independent variables present in the sample, the predicted probabilities \hat{y}_{gj} ($j=1,2,\dots,r$) will necessarily fall inside the unit interval. This is, of course, in itself a highly desirable feature of a probability model - to say the least. It does, however, also mean that the estimated slope parameters are quite sensitive to outliers.

Note also that for other strings of variable values x_h (say) than those present in the sample, there is no inherent guarantee that the predicted probabilities $\hat{y}_{hj} = x_h' \hat{\beta}_j$ ($j=1,2,\dots,r-1$) will obey $0 \leq \hat{y}_{hj} \leq 1$.

3. THE LOG-LINEAR MODEL

3.1. Introduction

Recall that we are dealing with a situation in which our sample consists of s vectors $(n_{g1}, n_{g2}, \dots, n_{gr})$ ($g=1,2,\dots,s$), each vector being multinomially distributed with parameters n_g and $(p_{g1}, p_{g2}, \dots, p_{gr})$. The probabilities p_{gj} are assumed to be known functions of some known attribute vector x_g and an unknown parameter vector β_j to be estimated. The linear model (4.12)-(4.15) is one possible way of specifying the functional relationship between p_{gj} and β_j . Another model in widespread use is the multinomial logit model:

$$(5.19) \quad \left\{ \begin{array}{l} p_{gj} = \frac{\exp(x_g' \beta_j)}{r-1 + \sum_{h=1}^{r-1} \exp(x_g' \beta_h)} \quad (j=1,2,\dots,r-1) \\ p_{gr} = \frac{1}{r-1 + \sum_{h=1}^{r-1} \exp(x_g' \beta_h)} \end{array} \right.$$

Plotting p_{gj} against $x_g' \beta_j$ we obtain a sigmoid-shaped curve, such that for all (finite) values of the argument, p_{gj} will belong to the (open) unit interval.

In (5.19), the dependent variables y_{gj} are viewed asymmetrically with respect to the entire vector of independent variables x_g . When all independent variables are qualitative, so that the data set can be collapsed into a contingency table, it is, however, common to treat all variables symmetrically - assumed dependent and independent ones alike. This approach is often referred to as the log-linear model. The multinomial logit model and the log-linear model are, however, essentially the same, in the sense that the probabilities p_{gj} in the former are obtained as the conditional probabilities calculable from the latter. That is, this correspondence is valid as long as the log-linear model contains a full set of interaction terms between all independent variables. For this reason the terms "multinomial logit" and "log-linear" models are used more or less indiscriminately in this paper.

3.2. The two-stage generalized least squares (2SGLS) estimator

Recall (5.19), perform the transformation

$$(5.24) \quad \log(p_{gj}/p_{gr}) = x_g' \beta_j \quad (g=1,2,\dots,s; j=1,2,\dots,r-1)$$

and define the error term

$$(5.25) \quad w_{gj} = \log(y_{gj}/y_{gr}) - \log(p_{gj}/p_{gr}).$$

y_{gj} denotes, as in section 2, the observed proportion of individuals with outcome j , among those with attribute vector x_g . Define

$$(5.26) \quad f_{gj} = \log(y_{gj}/y_{gr}) = x_g' \beta_j + w_{gj}$$

or, in vector notation,

$$(5.27) \quad f = Z\beta + w$$

by analogy to (4.20).

After suitable algebraic manipulation, it follows that

$$(5.29) \quad w_{gj} = (u_{gj}/p_{gj} - u_{gr}/p_{gr}) + v_{gj}$$

where the term

$$(5.30) \quad v_{gj} = \sum_{h=2}^{\infty} \frac{1}{h} (-1)^{h-1} \left[\left(\frac{u_{gj}}{p_{gj}} \right)^h - \left(\frac{u_{gr}}{p_{gr}} \right)^h \right]$$

is of order less than n_g^{-1} in probability. Thus, it does not affect the asymptotic properties of w_{gj} or f_{gj} . In large samples, one may assume

$$(5.31) \quad w_{gj} \approx u_{gj}/p_{gj} - u_{gr}/p_{gr}$$

and hence

$$(5.32) \quad E(w_{gj}) \approx 0$$

$$\begin{aligned}
 (5.33) \quad \text{var}(w_{gj}) &\approx p_{gj}^{-2} \text{var}(u_{gj}) + p_{gr}^{-2} \text{var}(u_{gr}) - 2p_{gj}^{-1}p_{gr}^{-1} \text{cov}(u_{gj}, u_{gr}) \\
 &= \frac{1-p_{gj}}{n_g p_{gj}} + \frac{1-p_{gr}}{n_g p_{gr}} + \frac{2}{n_g} \\
 &= n_g^{-1} (p_{gj}^{-1} + p_{gr}^{-1}) \quad (j=1,2,\dots,r-1)
 \end{aligned}$$

$$\begin{aligned}
 (5.34) \quad \text{cov}(w_{gj}, w_{gk}) &= E \left[w_{gj} \cdot w_{gk} \right] - E(w_{gj}) \cdot E(w_{gk}) \\
 &\approx E \left[\left(\frac{u_{gj}}{p_{gj}} - \frac{u_{gr}}{p_{gr}} \right) \left(\frac{u_{gk}}{p_{gk}} - \frac{u_{gr}}{p_{gr}} \right) \right] \\
 &= p_{gj}^{-1} p_{gk}^{-1} \text{cov}(u_{gj}, u_{gk}) - p_{gk}^{-1} p_{gr}^{-1} \text{cov}(u_{gk}, u_{gr}) \\
 &\quad - p_{gj}^{-1} p_{gr}^{-1} \text{cov}(u_{gj}, u_{gr}) + p_{gr}^{-2} \text{var}(u_{gr}) \\
 &= (n_g p_{gr})^{-1} \quad (g=1,2,\dots,s; j \neq k)
 \end{aligned}$$

$$(5.35) \quad \text{cov}(w_{gj}, w_{hk}) \approx 0 \quad (g \neq h).$$

Thus, there is an extremely simple and direct relationship between the (asymptotic) moments of w_{gj} and those of the "original" error terms u_{gj} :

$$\begin{aligned}
 (5.36) \quad \Sigma_W \equiv \text{Cov}(w) &\approx \begin{bmatrix} N^{-2} \Sigma^{11} & N^{-2} \Sigma^{12} & \dots & N^{-2} \Sigma^{1,r-1} \\ N^{-2} \Sigma^{21} & N^{-2} \Sigma^{22} & \dots & N^{-2} \Sigma^{2,r-1} \\ \vdots & \vdots & \ddots & \vdots \\ N^{-2} \Sigma^{r-1,1} & N^{-2} \Sigma^{r-1,2} & \dots & N^{-2} \Sigma^{r-1,r-1} \end{bmatrix} \\
 &= (I_{r-1} \otimes N^{-2}) \Sigma^{-1} = M^{-2} \Sigma^{-1}.
 \end{aligned}$$

Recall that Σ^{jj} and Σ^{jk} are the blocks of the partitioned inverse covariance matrix Σ^{-1} , as defined by (4.52)-(4.54). In the special case $N = I$ (non-replicated data), (5.36) reduces to

$$\Sigma_W = \Sigma^{-1}.$$

The 2SGLS estimator of β in the multinomial logit model is therefore (approximately) given by

$$(5.39) \quad \underline{d} = (Z' \hat{\Sigma}_W^{-1} Z)^{-1} Z' \hat{\Sigma}_W^{-1} \underline{f}.$$

Here, $\hat{\Sigma}_W^{-1}$ is the consistent estimator of Σ_W^{-1} formed by replacing p_{gj} by y_{gj} in the expression for Σ_W^{-1} ($= \Sigma M^2$).

The 2SGLS estimator \underline{d} (say) has asymptotic distribution given by

$$(5.40) \quad \sqrt{n} (\underline{d} - \underline{\beta}) \xrightarrow{L} N(0, \lim_{n \rightarrow \infty} n(Z' \hat{\Sigma}_W^{-1} Z)^{-1}).$$

The asymptotic covariance matrix is consistently estimated by

$$(5.41) \quad n \hat{\Sigma}_d = n(Z' \hat{\Sigma}_W^{-1} Z)^{-1} = n(Z' M^2 \hat{\Sigma} Z)^{-1}$$

whence it can be shown that under the null hypothesis $H_0: K\underline{\beta} = \underline{\kappa}$ (4.46):

$$(5.42) \quad n(K\underline{d} - \underline{\kappa})' [K \hat{\Sigma}_d K']^{-1} (K\underline{d} - \underline{\kappa}) \xrightarrow{L} \chi^2(\nu).$$

An asymptotic level α test of H_0 vs. $H_1: K\underline{\beta} \neq \underline{\kappa}$ consists in rejecting H_0 whenever the statistic (5.42) exceeds the $(1-\alpha)$ quantile of the χ^2 distribution with ν degrees of freedom, ν being the rank of K .

An IGLS estimator can be computed also for the multinomial logit model with replicated observations (grouped data). The procedure is analogous to the one described for the linear model. Here, one does not risk negative or larger-than-unity intermediate probability estimates.

In the dichotomous case ($r=2$), the 2SGLS estimator is algebraically identical to the well-known minimum logit chi-square estimator, proposed by Berkson (1944, 1953 and 1955a). This result can be viewed as analogous to Proposition 4.1, which relates the 2SGLS estimator in the linear model to the Neyman chi-square statistic.

3.3. The maximum likelihood (ML) estimator

Plugging (5.19) into (4.83) we obtain the log-likelihood function for the multinomial logit model

$$(5.44) \quad \log L = \sum_{g=1}^s \left\{ \sum_{h=1}^{r-1} n_{gh} x'_{g^2} \beta_h - n_g \log \left[1 + \sum_{h=1}^{r-1} \exp(x'_{g^2} \beta_h) \right] \right\} + \text{constant}.$$

By differentiation,

$$(5.45) \quad \frac{\partial \log L}{\partial \beta_{jk}} = \sum_{g=1}^s \left\{ n_{gj} x_{gk} - n_g x_{gk} \cdot \exp(x'_{g^2} \beta_j) \left[1 + \sum_{h=1}^{r-1} \exp(x'_{g^2} \beta_h) \right]^{-1} \right\} \\ = \sum_{g=1}^s \left\{ n_g x_{gk} (y_{gj} - p_{gj}) \right\}$$

$$(5.47) \quad \frac{\partial^2 \log L}{\partial \beta_{jk} \partial \beta_{ih}} = \sum_{g=1}^s \left[-n_g x_{gk} \frac{\partial p_{gj}}{\partial \beta_{ih}} \right] \\ = \sum_{g=1}^s \left[-n_g x_{gk} x_{gh} (\delta_{ij} p_{gi} - p_{gi} p_{gj}) \right],$$

where δ_{ij} is the Kronecker delta. In matrix notation, (5.47) becomes

$$(5.48) \quad \frac{\partial^2 \log L}{\partial \underline{\beta} \partial \underline{\beta}'} = -(I_{r-1} \otimes X')(I_{r-1} \otimes N^2) \Sigma (I_{r-1} \otimes X) \\ = -Z' M^2 \Sigma Z = -n Z' Q_n \Sigma_0 Z = -Z' \Sigma_W^{-1} Z.$$

The ML estimates are obtained by solving the non-linear system of equations

$$(5.46) \quad \frac{\partial \log L}{\partial \beta_{jk}} = 0 \quad (j=1,2,\dots, r-1; k=1,2,\dots, m).$$

The information matrix is obtained from the 2nd order conditions (5.48):

$$(5.49) \quad \Psi = -E \left[\frac{\partial^2 \log L}{\partial \underline{\beta} \partial \underline{\beta}'} \right] = Z' \Sigma_W^{-1} Z = n Z' Q_n \Sigma_0 Z.$$

A consistent estimator of the information matrix is

$$(5.50) \quad \tilde{\Psi} = Z' \tilde{\Sigma}_W^{-1} Z = Z' M^2 \tilde{\Sigma} Z,$$

where $\tilde{\Sigma}_W^{-1}$ and $\tilde{\Sigma}$ denote ML estimates of the matrices Σ_W^{-1} and Σ , respectively.

The ML estimator is asymptotically efficient. Thus

$$(5.51) \quad \sqrt{n} (\underline{e} - \underline{\beta}) \xrightarrow{L} N(0, \lim_{n \rightarrow \infty} n \Psi^{-1}).$$

and, under $H_0: K\underline{\beta} = \underline{\kappa}$ (4.46),

$$(5.52) \quad (K\underline{e} - \underline{\kappa})' \cdot [K \tilde{\Psi}^{-1} K']^{-1} \cdot (K\underline{e} - \underline{\kappa}) \xrightarrow{L} \chi^2(\nu).$$

The test statistic (5.52) has the same asymptotic distribution as the likelihood ratio chi-square statistic $-2 \log(L_0/L_1)$, where L_f denotes the likelihood calculated under H_f ($f = 0, 1$).

In the replicated observations case, the testing problem is conveniently viewed within the framework of the symmetric log-linear model. Here, the likelihood ratio chi-square statistic is usually denoted by G^2 . To be precise, the statistic for testing a hypothesis, say $H_g: \underline{\theta} \in \Omega_g$, about some parameter vector $\underline{\theta}$, is denoted by $G^2(H_g)$. This statistic is relevant when the set of admissible hypotheses contains the most general (saturated) model. If we want to test H_g conditionally on some other hypothesis $H_f: \underline{\theta} \in \Omega_f$, such that $\Omega_g \subset \Omega_f$, a more powerful test is obtained by using the χ^2 -statistic

$$(5.80) \quad G^2(H_g | H_f) = G^2(H_g) - G^2(H_f).$$

Goodman (1970 and 1971) demonstrates how this principle can be used to test sequentially a whole chain, or tree, of nested "hierarchical" hypotheses in the log-linear model. We revert to this question in section 5.7 of this summary.

4. A NOTE ON THE COEFFICIENT OF DETERMINATION (R^2) IN A LINEAR MODEL WITH BINARY RESPONSE

When confronting linear and log-linear models and inference methods, we need some criterion upon which to base the comparison. Various goodness-of-fit measures have been considered, few of which are, however, "neutral" with respect to the various probability models and/or estimation techniques. The OLS technique applied to the linear model leads, e.g., inherently to maximization of R^2 - the coefficient of determination, while 2SGLS is, as we have seen, equivalent to minimizing the Neyman χ^2 -statistic. ML applied to the log-linear model means minimizing G^2 - the likelihood ratio χ^2 -statistic.

In the process of judging the relevance of these diverse goodness-of-fit measures, we have been led to ask in somewhat greater detail what are the properties of the well-known R^2 statistic for a linear model with binary dependent variable. It stands to reason that in such cases R^2 will usually turn out low, since the error variance is abnormally high compared to the expected value of the response variable. This is, at least, the case when the dummy dependent variables are realizations of a Bernoulli process with true probability in the neighbourhood of 0.5.

Indeed, empirical work done with this kind of model often does come out with unusually low R^2 measures. While perhaps not the best statistical technique for analyzing categorical data, the linear regression model with OLS estimation is in wide-spread use among social scientists. We have therefore set out to investigate just how low an R^2 can reasonably be expected, under varying conditions.

In the replicated observations case (i.e., grouped data), when the coefficient vector is estimated using formula (4.66), the relevant formula for R^2 is

$$(7.5) \quad R^2 = 1 - \frac{\hat{u}'M\hat{u}}{y'M^2Ay}$$

Here, \hat{u} denotes the $((tx)1)$, say) vector of residuals and

$$(7.3) \quad A = (I_t - \frac{1}{t} \underline{j}_t \underline{j}_t')$$

is an idempotent matrix transforming a vector of observations to deviations from their mean. We denote by \underline{j}_t , or just \underline{j} , a $(tx)1$ vector of unit elements.

Formula (7.5) can be rewritten as

$$(7.12) \quad R^2 = \frac{\hat{y}'M\hat{y} - \frac{1}{s(r-1)} (\underline{j}'M^{\frac{1}{2}}\hat{y})^2}{y'My - \frac{1}{s(r-1)} (\underline{j}'M^{\frac{1}{2}}y)^2},$$

drawing on the notation from section 2 of this summary. Furthermore, recalling formulae (4.34)-(4.37), it can be shown for the dichotomous case ($r=2$) that

$$(7.34) \quad \text{plim } R^2 = \frac{\text{plim } \frac{1}{n} \hat{y}'M\hat{y} - \frac{1}{s} \text{plim } \frac{1}{n} (\underline{j}'M^{\frac{1}{2}}\hat{y})^2}{\text{plim } \frac{1}{n} y'My - \frac{1}{s} \text{plim } \frac{1}{n} (\underline{j}'M^{\frac{1}{2}}y)^2}$$

$$= \frac{p'Qp - \frac{1}{s} (\underline{j}'Q^{\frac{1}{2}}p)^2}{p'Qp - \frac{1}{s} (\underline{j}'Q^{\frac{1}{2}}p)^2} = 1.$$

When the number of observations in each group approaches infinity, the error variance vanishes and the fit becomes perfect - an intuitively obvious result. (It is, however, based on the rather crucial assumption that the probability model used is the "true" one.)

Now, let us turn to the more interesting case where each individual in the (same) sample is counted as one separate observation, i.e. the model is formulated as (4.57). When running an OLS regression on this model, the coefficient of determination will be computed using the formula

$$(7.13) \quad R^{*2} = 1 - \frac{\hat{u}^{*'}\hat{u}^*}{y^{*'}Ay^*}$$

or, equivalently,

$$(7.22) \quad R^{*2} = \frac{\hat{y}'My - \frac{1}{n(r-1)} (i' My)^2}{\tilde{i}' My - \frac{1}{n(r-1)} (i' My)^2}.$$

As before, the asterisk is meant to indicate that we are working with a string of individual (non-replicated) observations.

The statistic (7.22) obeys

$$(7.33) \quad \begin{aligned} \text{plim } R^{*2} &= \frac{\text{plim } \frac{1}{n} \hat{y}' My - \text{plim} \left(\frac{1}{n} i' My \right)^2}{\text{plim } \frac{1}{n} \tilde{i}' My - \text{plim} \left(\frac{1}{n} i' My \right)^2} \\ &= \frac{p' Qp - (i' Qp)^2}{\tilde{i}' Qp - (i' Qp)^2} < 1. \end{aligned}$$

As emphasized in section 2.3, both ways of organizing and analyzing the data (to group or not to group - that's the question!) yield identical parameter estimates. The goodness-of-fit measures emanating from either procedure are, however, quite different. The R^{*2} statistic does not approach unity when the sample size increases. From comparing formulae (7.12) and (7.22) it is readily seen that, because $y_{gj}^2 < y_{gj}$ for $0 < y_{gj} < 1$,

$$(7.25) \quad R^{*2} < R^2$$

even in finite samples.

Just how different are R^{*2} and R^2 ? For the sake of the argument, let us make the simplifying assumption $n_g = n/s$, i.e.

$$(7.36) \quad M = \frac{n}{s} \cdot I.$$

The sample is evenly distributed across all s groups in the marginal table of independent variables. Also, from here on we will be considering only the dichotomous case ($r=2$).

Now the following simple formula can be derived:

$$\frac{R^{*2}}{R^2} = \frac{\hat{\tau}_1}{\hat{\tau}_2} < 1,$$

where

$$(7.40) \quad \hat{\tau}_1 = s \tilde{y}' y - (i' y)^2$$

$$(7.41) \quad \hat{\tau}_2 = s i' y - (i' y)^2.$$

Expanding $E[\hat{\tau}_1/\hat{\tau}_2]$ into a Taylor series and retaining terms up to and including the 2nd power, it can be shown that

$$(7.47) \quad E(\hat{\tau}_1/\hat{\tau}_2) \approx \frac{\tau_1}{\tau_2} + \frac{\tau_1}{\tau_2} \left[\frac{1}{\tau_2} \text{var}(\hat{\tau}_2) - \frac{1}{\tau_1 \tau_2} \text{cov}(\hat{\tau}_1, \hat{\tau}_2) \right].$$

where τ_1 and τ_2 are the expected values of $\hat{\tau}_1$ and $\hat{\tau}_2$, respectively.

A rough indication of the expected ratio of R^{*2} to R^2 can be obtained by disregarding the bracketed terms of (7.47).

Define

$$(7.53) \quad \mu = \frac{1}{s} \sum_{g=1}^s p_{g1}$$

$$(7.54) \quad \sigma^2 = \frac{1}{s} \sum_{g=1}^s p_{g1}^2 - \mu^2$$

and

$$(7.55) \quad \omega = \mu(1-\mu).$$

After some tedious algebra we derive the following simple formulae:

$$(7.56) \quad \frac{\tau_1}{\tau_2} = \frac{(n-s+1)\sigma^2 + (s-1)\omega}{\sigma^2 + (n-1)\omega}$$

$$(7.57) \quad \lim_{n \rightarrow \infty} (\tau_1/\tau_2) = \sigma^2/\omega.$$

τ_1/τ_2 depends on the true probabilities p_{11}, \dots, p_{s1} only through the two summary measures σ^2 and ω . σ^2 measures the spread, while ω indicates how close to 0.5 is the average of the true probabilities.

Upon differentiation of (7.56) it is seen that the ratio τ_1/τ_2 increases with σ^2 and decreases with ω . $\text{plim } R^{*2}$ coincides, of course, with $\lim_{n \rightarrow \infty} \tau_1/\tau_2$.

The ratio τ_1/τ_2 and the value of $\text{plim } R^{*2}$ have been calculated for a number of numerical examples on p_{11}, \dots, p_{s1} . The examples are explained in table 7.1 (of the Norwegian text) and in figure 7.1 (for $s=24$). A programmable pocket calculator was used to perform the calculations. Results are given in tables 7.2 - 7.6.

When the spread of the true probabilities is small and/or their average is close to one half, the coefficient of determination will often turn out extremely low, no matter how correctly the model has been specified. The most striking of our numerical examples is no. 6, where all the true probabilities are clustered at one end of the unit interval. In this case R^{*2} - the proportion of "explained variance" - cannot under any circumstances be expected to exceed 5 per cent.

5. EMPIRICAL STUDIES

5.1. *The data set*

In order to compare the relative performance of the linear (as estimated by OLS or 2SGLS) and the log-linear (ML) models, a data set taken from the Norwegian labour force sample surveys in 1976 was used. The full data set is exhibited in table 8.1. It consists of 9 523 male respondents aged 16-74 and forms a 3 x 3 x 3 x 5 contingency table. The response variable is taken to be (weekly) "number of hours worked" (T), while the independent variables are "age" (A), "level of education" (U) and "marital status" (S). Thus, $n=9523$, $r=3$ and $s=45$.

The great bulk of the respondents belong to groups (g) where the proportion of persons employed full-time is between 85 and 90 per cent, cf. figure 8.1. Only very few respondents belong to groups with average response frequencies (y_{gj}) higher than 0.95 or lower than 0.05.

5.2. *Empty cells*

The data set contains 23 empty cells. If the 2SGLS procedure is to be practicable, a (small) constant δ (say) must be added to each cell. Following Haldane (1955) and Anscombe (1956) we have tried $\delta = 0.5$ and also, in order to minimize the arbitrary element, $\delta = 0.01$.

5.3. *Method of calculation*

The ML estimates of the log-linear model were calculated using the ECTA program. The OLS and 2SGLS estimators of the linear model were programmed in DATSY - a general system for (i.a.) matrix manipulation developed in the Central Bureau of Statistics.

5.4. *Comparing goodness-of-fit*

Some summary results of the experiments are given in tables 9.1 and 9.2. Four different probability models have been tried: (1) the linear model without interaction terms, (2) the linear model with all 1st-order interaction terms, (3) the multinomial logit model without interaction terms, and (4) the multinomial logit model with all 1st-order interaction terms. Models (1) and (2) have been estimated both by OLS and 2SGLS, while only ML has been used to estimate models (3) and (4). For each of the six procedures the following goodness-of-fit measures have been calculated: R^2 , χ_N^2 , G^2 , and C, the last one being defined by

$$(7.65) \quad C = \frac{1}{2} \sum_{g=1}^s \sum_{j=1}^r n_g |y_{gj} - \hat{y}_{gj}| = \sum_{\{\hat{u}_{gj} > 0\}} \sum_g n_g (y_{gj} - \hat{y}_{gj}).$$

R^2 is the coefficient of determination (7.5), χ_N^2 is the Neyman chi-square (4.91), and G^2 is the likelihood ratio chi-square statistic, which can be written

$$G^2 = 2 \sum_{g=1}^s \sum_{j=1}^r n_g y_{gj} \log(y_{gj} / \hat{p}_{gj}).$$

OLS applied to the linear model leads intrinsically to maximization of R^2 , while 2SGLS is equivalent to $\min \chi_N^2$. In the log-linear model, ML means minimizing G^2 . Thus, out of our four goodness-of-fit measures, only C is in a sense "neutral" with respect to all three estimation techniques. C measures the amount of "misclassification" and is used, i.a., by Domencich and McFadden (1975, p. 124-125). The quantity $100 \cdot C/n$ tells us what percentage of the entire sample is "misclassified". $100 \cdot \frac{2}{rn} \cdot C$ is the (weighted) average of all absolute residuals $|\hat{u}_{gj}|$, measured in percentage points. The stochastic properties of C are largely unknown.

Examining table 9.1 we note that when $\delta = 0.01$, the 2SGLS procedure does quite badly. The 23 empty cells are, in fact, allowed to govern the entire estimation process through the arbitrary constant δ . Indeed, from formula (4.91) it is readily seen that when y_{gj} is very small, cell (g,j) will obtain a lot of weight.

Table 9.2 exhibits the results for $\delta = 0.5$. These are more "normal". As suspected, the "non-neutral" goodness-of-fit measures do not yield conclusive evidence as to the relative merits of the different techniques - they contradict each other. As measured by C, however, OLS/linear model misclassifies 3.74 per cent, for 2SGLS/linear model the figure is 3.54 per cent, while for ML/log-linear model it is 3.20 per cent. These results apply when the true model contains no interaction terms. In the model with all 1st-order interaction terms the differences between the three techniques are, not surprisingly, negligible.

When interpreting these results, the structure of the data set, as shown by figure 8.1, should be kept in mind. In a data set where the y_{gj} were scattered more evenly along the unit interval, and/or more of them were close to 0 or 1, the superiority of the log-linear model would probably be more pronounced.

Another warning is in order. Comparing goodness-of-fit measures should not replace a priori thinking. Ideally, the choice of probability model should be made so as to conform with "reality" - the "true" stochastic process having generated the data. It is conceivable that, in a particular random sample, the "true" model is not the one which scores best on some arbitrary goodness-of-fit test. Only in the absence of a priori guidelines can goodness-of-fit measures give us a clue as to which model to prefer.

5.5. Comparing the variance of OLS and 2SGLS

A consistent estimator of the covariance matrix of the 2SGLS estimator in the linear model is given by (4.45). Similarly, a consistent estimator of the OLS covariance matrix is obtained by substituting y_{gj} for p_{gj} in expression (4.69). Let $\widehat{\text{var}}(b_{jk})$ and $\widehat{\text{var}}(\hat{\beta}_{jk})$ denote the diagonal elements of these matrices, respectively, and define the relative efficiency measure

$$(9.1) \quad \hat{\epsilon}(b_{jk}, \hat{\beta}_{jk}) = [\widehat{\text{var}}(\hat{\beta}_{jk}) / \widehat{\text{var}}(b_{jk})]^{-1/2} \quad (j=1,2,\dots, r-1, k=1,2,\dots, m).$$

The values of $\hat{\epsilon}(b_{jk}, \hat{\beta}_{jk})$ are given in table 9.3 for the model without interaction terms and $\delta = 0.5$. The estimated gain in efficiency by going from OLS to 2SGLS varies from 1.4 to about 18 per cent, as measured by the standard deviation.

In the model with all 1st-order interaction terms the difference in efficiency between the two methods is much smaller, cf. table 9.4.

Often our primary interest is focused not on the single parameter values β_{jk} , but rather on some particular contrasts $\beta_{ji} - \beta_{jk}$ ($i \neq k$). The precision with which one estimates such contrasts depends on the variances of the two parameter estimators as well as on their covariance. Relative efficiency estimates for the contrasts of most interest are shown in table 9.5, for the model without interaction terms. The picture turns out very much the same as for the single parameter estimates, as shown in table 9.3.

Note, however, that out of the two methods, only OLS is unbiased. Thus, one ought to compare the mean square errors rather than the variances.

5.6. Efficiency gain with respect to the saturated model

Assuming that the linear model without interaction terms is the true model, one can estimate how much precision is gained from estimating the p_{gj} ($g=1,2,\dots,s, j=1,2,\dots,r$) by 2SGLS rather than using the observed cell frequencies y_{gj} , i.e. the saturated model. This is shown in figure 9.1.

As would be expected, the gain in efficiency decreases rapidly with increasing n_g . Using a non-saturated probability model amounts to extracting information not only from the one cell whose true p_{gj} is being estimated, but from adjacent cells as well. The potential gain from this is, of course, largest for those cells which themselves contain relatively little information (i.e., n_g small).

5.7. Testing hypotheses in the log-linear model

Tables 9.6-9.8 show the results of our efforts to test a tree of hierarchical hypotheses in the log-linear model. The testing results appear to be sensitive to the value of δ - the arbitrary constant added to avoid empty cells.

The (symmetric) model without 4-factor effects (H_1) is accepted (i.e., no 2nd-order interaction terms in the asymmetric formulation). Three hypotheses implying absence of 3-factor-effects are tested conditionally on H_1 , of which only one (H_4) is rejected at the 1 per cent level. The two remaining hypotheses (H_2 and H_3) can, however, not be retained simultaneously: $H_5 = H_2 \cap H_3$ is rejected when tested against H_1 . Out of H_2 and H_3 , H_2 yields the higher p-value, thus this model is the one finally retained at the 1 per cent level.

We conclude that there are no 3-factor-effects between hours worked (T), marital status (S) and level of education (U). All other 3-factor-effects are, however, present.

5.8. Testing hypotheses in the linear model

Recalling proposition 4.2 (section 2.2 of this summary) a completely analogous procedure may be followed to test for absence of interaction effects in the linear model. Here, we use the Neyman chi-square statistic. The results of these test are exposed in table 9.9. The hypotheses considered (H_1 to H_5) are the "same" as those contemplated for the log-linear model, in the sense that corresponding β -parameters are set equal to zero in both models (cf. (4.20) and (5.27)).

The p-values for testing corresponding hypotheses are generally lower in the linear than in the log-linear model (see table 9.10). However, in both models the same hypothesis (H_2) is retained finally, using a 1 per cent level of significance.

6. Synthesis

This paper is concerned with a situation in which the true probabilities (p_{gj}) of several independent, multinomially distributed qualitative response variables are assumed to depend on the same set of known independent variables (x_{gk}) and unknown parameters (β_{jk}). A number of different ways to specify the functional relationship between the multinomial probabilities p_{gj} and the independent variables are studied, as well as inference methods relevant to each model specification. The arsenal of available inference techniques varies with the model specification and, above all, with the level of measurement of the independent variables. Figure 10.2 summarizes what techniques are practicable in certain selected situations.

The choice of probability model for a qualitative response variable can be viewed from at least three different angles. One can depart from a (say) "purist" attitude, a "pragmatic" attitude, or an "economic" attitude.

According to the purist attitude, a probability model must satisfy certain requirements as to inner consistency and logic. Moreover, the purist believes that there exists a "true" model, or, at the very least, that out of two incompatible, competing models, one can often be designated as superior to the other on a priori grounds. To the purist, any model which can yield "probability" estimates outside the unit interval is an anomaly. The linear model truncated at $p = 0$ and $p = 1$ will, on the other hand, usually be considered as unreasonable due to its (arbitrary) discontinuities.

Advocates of the pragmatic view contend that only the practical results count. In the social sciences, data originate seldom as a result of some known, well-defined stochastic process. Rather, the purpose of probability models is to help us organize and condense a vast and otherwise intractable body of information, so as to illuminate a few general phenomena. We choose that statistical technique which solves this task most efficiently. A model allowing negative or larger-than-unity probability estimates is inefficient only in the sense that it disregards the information that such instances cannot occur, and only to the extent that this reduces the reliability of the results.

The economic attitude consists in choosing the least expensive technique, other things being largely equal.

Figure 10.1 compares the linear, logit and probit binary models. The scale is in a sense the same for all three curves, σ being the standard deviation corresponding to each quantal response curve considered as a cumulative distribution function. (The linear model is assumed truncated at either end of the unit interval, i.e. the distribution is rectangular between 0 and 1.)

The probit and logit models differ only marginally. The former has somewhat slimmer tails and is thus slightly less robust with respect to outliers. Other authors (i.a., Goldfeld and Quandt (1972)) have found that the sampling properties of the ML probit estimators are rather unsatisfactory and that numerical convergence to the ML-estimator is often slow. Moreover, the extension from the binomial to the multinomial probit model is not completely trivial. These are precisely the reasons why the probit model is not studied in any greater detail here - we prefer the logit model. It is, of course, conceivable to encounter situations where probit is the a priori correct model. However, we believe that in the social sciences the purist attitude to statistical method is often not the most fruitful.

Although nothing can be said in general about the relative merits of the linear and logit models, in most cases the logit model seems preferable from a purist point of view. A pragmatic appraisal also seems to favour the logit model, although a universal criterion for judging one model against the other can hardly be found. Our index of misclassification (C) is the relevant measure of goodness as long as the cost of misclassification is the same for all respondents and for every type of error.

Only economic considerations will usually favour the linear model as against logit. This is at least the case when the independent variables are quantitative (non-replicate data).

In the linear model, the 2SGLS estimator, where practicable, is superior to OLS. Suitable ad hoc corrections must, however, be made to avoid empty cells.

In the multinomial logit model with replicate data, the 2SGLS estimator is at least as good as the ML estimator. When grouping is necessary prior to the 2SGLS procedure, this estimator can still compete with ML in small samples, provided the true probabilities are not too close to 0 or 1 (Berkson (1955a), Domencich and McFadden (1975), Berkson (1980), Amemiya (1980)).

As applied to our data set, the ML/logit, 2SGLS/linear and OLS/linear techniques do not yield widely differing results as measured by the misclassification index. This may be partly due to the absence of frequencies close to 0 or 1 in our data set.

A. GRUNNBEGREPER OG DEFINISJONER

A.1. Notasjon

Matriser er i dette notatet skrevet med store greske eller latinske bokstaver (f.eks. X, Σ). Apostrof (') betegner transponering.

Vektorer skrives med små greske eller latinske bokstaver med krøll under (f.eks. x_g, β). Bokstaver uten apostrof betegner søylevektorer, bokstaver med apostrof betegner rekkevektorer.

Skalarer skrives med små greske eller latinske bokstaver uten krøll.

Unntak fra disse reglene har vi gjort for en del føyningsmål, der en bestemt notasjon er innarbeidet i litteraturen (f.eks. R^2, G^2, L m.fl.), likeledes for den normale kumulative fordelingsfunksjonen (Φ) og for sannsynligheter ($P(y_i = 1)$). Mengder, hypoteser og i noen tilfeller grupperingsveier vil også bli skrevet med stor bokstav.

Vi skiller ikke mellom stokastiske variable og de observerte verdier av disse variable. Dette skulle ikke gi opphav til forvirring.

Følgende standardtegn vil dessuten bli brukt:

I_m = identitetsmatrisen av dimensjon m . Ofte vil vi skrive bare I istedenfor I_m .

$\underline{j}_m = (1, 1, \dots, 1)'$, en søylevektor bestående av m ett-tall. Ofte vil vi skrive bare \underline{j} istedenfor \underline{j}_m . Summen av alle elementene i en vektor \underline{a} bare dermed skrives som $\underline{j}'\underline{a}$ eller $\underline{a}'\underline{j}$.

\emptyset_{mn} = en $(m \times n)$ matrise bestående av bare nuller. Ofte vil vi skrive bare \emptyset istedenfor \emptyset_{mn} .

$\underline{0}_m = (0, 0, \dots, 0)'$, en søylevektor bestående av m nuller. Ofte vil vi skrive bare $\underline{0}$ istedenfor $\underline{0}_m$.

Elementet i i -te rad og j -te kolonne i matrisen A (henh. X, Λ e.l.) vil bli betegnet med: a_{ij} (henh. x_{ij}, λ_{ij} osv.). Hele matrisen A vil vi av og til skrive som $[a_{ij}]$, for å markere hvilket tall som står i posisjon (i, j) .

Symbolet $x \xrightarrow{L} y$ betyr "x konvergerer i fordeling mot y". Symbolet $x \xrightarrow{P} y$ betyr "x konvergerer i sannsynlighet mot y".

Skrivemåten $\text{plim } x = y$ blir brukt synonymt med $x \xrightarrow{P} y$. Betydningen av disse begrepene blir nærmere forklart i avsnitt A.3.

Symbolet δ_{ij} står for det såkalte "Kronecker delta", definert ved

$$(A.1) \quad \delta_{ij} = \begin{cases} 1 & \text{hvis } i = j \\ 0 & \text{hvis } i \neq j \end{cases}$$

Symbolet \otimes står for Kronecker-produkt, definert som følger. La A og B være to matriser der de enkelte elementer benevnes a_{ij} og b_{ij} , henholdsvis. A har dimensjon $(m \times n)$. Kronecker-produktet $A \otimes B$ er da den blokk-oppdelte matrisen

$$(A.2) \quad A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ a_{21}B & a_{22}B & \dots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{bmatrix}$$

Symbolet $\exp(x)$ blir brukt for eksponentialfunksjonen e^x , og symbolet $\log(x)$ eller $\log x$ for den naturlige logaritmen til x .

Skrivemåten $P(A)$ betyr "sannsynligheten for at hendelsen A inntreffer". For eksempel vil $P(x > y)$ bety "sannsynligheten for at x er større enn y ". $P(A|B)$ vil bety "den betingede sannsynlighet for A , gitt B ".

Skrivemåten

$$(A.3) \quad x \sim N(\mu, \sigma^2)$$

betyr " x er normalfordelt med forventning μ og varians σ^2 ". Skrivemåten

$$(A.4) \quad \underline{x} \sim N(\underline{\mu}, \Sigma)$$

betyr " \underline{x} er multinormalt fordelt med forventning $\underline{\mu}$ og kovariansmatrise Σ ". Skrivemåten

$$(A.5) \quad \underline{x} \sim \chi^2(\nu)$$

betyr " x er kji-kvadrat-fordelt med ν frihetsgrader".

Med $E(x)$ mener vi forventningen til x , $\text{var}(x)$ står for variansen til x , og $\text{cov}(x, y)$ står for kovariansen mellom x og y . $\text{Cov}(\underline{y})$, eller Σ_y , vil stå for kovariansmatrisen til \underline{y} . I noen tilfelle vil vi skrive $E_{\theta}(x)$ og $\text{var}_{\theta}(x)$ for å markere at momentene avhenger funksjonelt av parameteren θ .

A.2. Målenivå og variabeltyper

Det finnes en rekke typologier for klassifisering av variable. Den vanligste typologien innenfor sosiologien ble utarbeidet av Stevens (1946) og inndeles variablene etter deres målenivå: nominalnivå, ordinalnivå, intervallnivå og forholdstallsnivå.

Nominalnivå er det laveste målenivået i Stevens' typologi. Hver variabel kan anta et endelig antall gjensidig utelukkende og uttømmende verdier, men hver verdi tjener kun som en merkelapp eller navn (derav: nominalnivå) på vedkommende kategori. Det er ingen naturlig rangordning mellom verdiene. Vi kan ikke si at én verdi er "større" enn en annen, og følgelig er det heller ikke mulig å snakke om avstand mellom de ulike verdier. Eksempler på slike variable kan være kjønn, bosted, næring eller farge.

Ordinalnivå er det neste målenivået. Her finnes det en naturlig måte å rangordne verdiene på, og det gir mening å snakke om at én verdi er større enn en annen. Det gir imidlertid ingen mening å snakke om avstand mellom verdiene, f.eks. at verdi 3 er like mye større enn verdi 2 som verdi 2 er større enn verdi 1. Eksempel på en slik variabel kan være utdanningsnivå. Et annet eksempel her er nettopp Stevens' målenivåskala (nominal-, ordinal-, intervall- eller forholdstallsnivå)!

På intervallnivå, derimot, har det mening å snakke om avstand (intervall) mellom verdiene. Det klassiske eksempel er her temperaturskalaen, målt f.eks. i celsius- eller Fahrenheitgrader. Vi kan si at forskjellen mellom 0° og 10° C er like stor som mellom 10° og 20° C, men det gir ikke mening å si at 20° er "dobbelte så varmt" som 10° C. Nullpunktet i skalaen er nemlig vilkårlig valgt.

Det høyeste målenivået er forholdstallsnivå. For slike variable har det mening å definere forholdet mellom to verdier: en person på 80 år er dobbelt så gammel som en på 40. En arbeidsledighetsprosent på 2,4 er dobbelt så høy som en på 1,2.

Det er mulig å redusere målenivået for en variabel målt på ordinalnivå eller høyere. Ved å se bort fra rangordningen på en ordinalskala får vi nominalnivå. Ved å gruppere verdiene på en forholdstallsvariabel får vi ordinalnivå. En slik prosess innebærer at vi kaster bort den informasjon som kjennetegner det høyere målenivået.

I denne rapporten vil skillet mellom nominal- og ordinalvariable ha liten betydning. Med et fellesnavn kan vi kalle disse variabeltypene for kvalitative variable. Heller ikke vil vi skille mellom intervall- og forholdstallsvariable, som vi med et fellesnavn kan kalle kvantitative variable.

En kvalitativ variabel med bare to mulige verdier kaller vi en dikotom variabel. Kjønn er en slik variabel. Når det ene utfallet tilordnes verdien 0 og det andre 1, vil vi snakke om binære variable, eller dummy-variable.

En kvalitativ variabel med flere enn to mulige utfall vil vi kalle polytom. En polytom variabel med r mulige verdier kan beskrives entydig ved hjelp av en vektor av $r-1$ binær-variable. En slik vektor vil vi kalle en multinær vektor. Variabelen alder i tabell 3.1 er en slik polytom variabel, representert ved en multinær vektor.

En variabeltypologi som i prinsippet går på tvers av Stevens' typologi er inndelingen i diskrete og kontinuerlige variable. Med en diskret variabel mener vi her en variabel som bare kan anta et tellbart antall verdier, det vil si at det går an å nummerere alle de mulige verdier av variabelen, slik at vi for enhver vilkårlig verdi kan si hvilket nummer i rekken denne er. Heltallsvariable er det vanligste eksemplet. For diskrete variable er det i alminnelighet en positiv sannsynlighet for at to stokastisk uavhengige forsøk kan gi samme verdi. Denne sannsynligheten er gjerne større jo færre mulige verdier variablene kan anta. Binærvariable er et spesialtilfelle med kun 2 mulige utfall. En kontinuerlig variabel, derimot, kan anta en hvilken som helst verdi innenfor et visst intervall på den reelle tallinje. Intervallet kan eventuelt gå fra minus til pluss uendelig. Dersom målingene skjer med uendelig høy nøyaktighet, er sannsynligheten null for at to uavhengige målinger kan gi samme resultat. I praksis skjer målingene med et begrenset antall gjeldende sifre, og det er da en positiv, om enn som oftest relativt liten sannsynlighet for sammenfallende utfall. Betydningen av dette forholdet blir klart i kapitlene 4 og 5 i denne artikkelen.

I praksis vil en ofte oppleve at denne inndelingen faller sammen med skillet mellom kvalitative og kvantitative variable. Kontinuerlige variable er nesten alltid målt på forholdstallsnivå, mens diskrete variable ofte er målt på nominal- eller ordinalnivå.

I en del av litteraturen vil en se uttrykket diskrete variable brukt som synonym for kvalitative variable. Et annet vanlig synonym er kategoriske variable.

En nærmere drøfting av spørsmålet om målenivå og variabeltyper kan en finne hos Hellevik (1971), Blaloch (1972), og dessuten hos Nie et al. (1975).

A.3. Elementer av lineær algebra og matematisk statistikk

I dette avsnittet refererer vi en del begreper og resultater som er nyttet under utledningen i deler av denne rapporten. Framstillingen bygger i hovedsak på Bickel og Doksum (1977) og Dhrymes (1970).

A.3.1. Definitte matriser

La A være en kvadratisk matrise. Dersom, for enhver vektor $x \neq 0$, vi har

$$(A.6) \quad x'Ax > 0$$

sies matrisen A å være positiv definitt. Dersom (A.6) gjelder som uekte ulikhet ($x'Ax \geq 0$) sies A å være ikke-negativ definitt eller positiv semidefinitt.

En nødvendig betingelse for at en matrise skal være positiv (ikke-negativ) definitt er at alle diagonalelementene er positive (ikke-negative). Enhver positiv definitt matrise er ikke-singulær.

La A og B være to kvadratiske matriser av samme dimensjon. Dersom matrisen

$$(A.7) \quad C = A - B$$

er positiv (ikke-negativ) definitt, vil vi si at A er større enn (minst like stor som) B . Dersom derimot matrisen

$$(A.8) \quad -C = B - A$$

er positiv (ikke-negativ) definitt, vil vi si at A er mindre enn (høyst like stor som) B .

Begrepene positive og ikke-negative tall (skalarer) er spesialtilfeller av de positiv, henholdsvis ikke-negativ definitte matriser.

A.3.2. Kovariansmatriser

En stokastisk vektor \underline{y} er en vektor der elementene y_1, y_2, \dots, y_r er stokastiske variable med en bestemt, simultan fordeling. Med kovariansmatrisen til \underline{y} , Σ_y , forstår vi matrisen av varianser og kovarianser mellom elementene y_1, y_2, \dots, y_r :

$$(A.9) \quad \Sigma_y = \text{Cov}(\underline{y}) = \begin{bmatrix} \text{var}(y_1) & \text{cov}(y_1, y_2) & \dots & \text{cov}(y_1, y_r) \\ \text{cov}(y_2, y_1) & \text{var}(y_2) & \dots & \text{cov}(y_2, y_r) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(y_r, y_1) & \text{cov}(y_r, y_2) & \dots & \text{var}(y_r) \end{bmatrix}$$

Siden $\text{cov}(y_i, y_j) = \text{cov}(y_j, y_i)$, er en kovariansmatrise alltid symmetrisk. Enhver kovariansmatrise er dessuten ikke-negativ definitt.

Definer nå den stokastiske vektoren

$$(A.10) \quad \underline{x} = A\underline{y}$$

der A er en vilkårlig ikke-stokastisk matrise slik at produktet $A\underline{y}$ eksisterer. Vi sier at \underline{y} er en lineærtransformasjon av \underline{y} , eller at \underline{x} er lineær i \underline{y} . Da er

$$(A.11) \quad \Sigma_x = \text{Cov}(\underline{x}) = A\Sigma_y A'.$$

La oss betrakte spesialtilfellet der A består av bare én rad, \underline{a}' . Produktet

$$(A.12) \quad x = \underline{a}'\underline{y}$$

er da en skalar, og vi sier at x er en lineærkombinasjon av \underline{y} . "Kovariansmatrisen" til $\underline{a}'\underline{y}$ degenererer til bare ett element:

$$(A.13) \quad \text{var}(x) = \underline{a}'\Sigma_y \underline{a}.$$

Dersom det finnes en vektor $\underline{a} \neq \underline{0}$ slik at $\underline{a}'\underline{y}$ blir en konstant, dvs. at $\text{var}(x)$ blir lik null, er kovariansematrisen Σ_y singular. Den er ikke-negativ definitt, men ikke positiv definitt. Den simultane fordelingen til \underline{y} sies da også å være singular. Dette innebærer at minst ett av elementene y_i ($i=1, \dots, r$) er identisk gitt som en lineærkombinasjon av de øvrige $r-1$ elementer.

Dersom imidlertid fordelingen til \underline{y} ikke er singular, vet vi at kovariansmatrisen Σ_y er positiv definitt.

A.3.3. Den multinormale fordeling

En r -dimensjonal stokastisk vektor \underline{x} sies å være multinormalt fordelt med forventning $\underline{\mu}$ og ikke-singular kovariansmatrise Σ ($\underline{x} \sim N(\underline{\mu}, \Sigma)$) dersom den simultane sannsynlighetstettheten til \underline{x} er gitt ved

$$(A.14) \quad f(\underline{x}) = \frac{1}{(2\pi)^{\frac{r}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu}) \right\}$$

La A være en $(v \times r)$ konstant matrise med rang v og \underline{a} en $(v \times 1)$ konstant vektor. En sentral egenskap ved multinormalfordelingen er da at

$$(A.15) \quad \underline{x} \sim N(\underline{\mu}, \Sigma) \Rightarrow A\underline{x} + \underline{a} \sim N(A\underline{\mu} + \underline{a}, A\Sigma A').$$

Dersom \underline{x} er multinormalfordelt er altså enhver transformasjon av formen $A\underline{x} + \underline{a}$ også multinormalfordelt. Videre kan en vise

$$(A.16) \quad \underline{x} \sim N(\underline{\mu}, \Sigma) \Rightarrow (\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu}) \sim \chi^2(r),$$

altså at den veide kvadratsummen $(\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu})$ er kji-kvadrat-fordelt med r frihetsgrader.

En grei referanse her er Scheffé (1959), s. 416-418.

A.3.4. Konvergens i sannsynlighet

La x_1, x_2, \dots være en følge av stokastiske variable. Vi sier at x_n konvergerer mot x i sannsynlighet dersom

$$(A.17) \quad \lim_{n \rightarrow \infty} P(|x_n - x| \geq \epsilon) = 0$$

for enhver $\epsilon > 0$. Intuitivt betyr dette at vi, ved å velge n stor nok, kan få x_n til å være så nær x vi bare vil, med en "til visshet grensende sannsynlighet".

Vi skriver da

$$(A.18) \quad x_n \xrightarrow{P} x$$

eller

$$(A.19) \quad \text{plim } x_n = x.$$

plim x_n står for "sannsynlighetsgrensen til x_n ".

Dersom x_n er definert som (det aritmetiske) gjennomsnittet av n uavhengige, identisk fordelte stokastiske variable med forventning μ og varians $\sigma^2 < \infty$, kan en vise at

$$(A.20) \quad x_n \xrightarrow{P} \mu.$$

Dette resultatet er kjent som de store talls lov ("weak law of large numbers").

Dersom en vilkårlig observator x_n konvergerer i sannsynlighet mot (den konstante) parameteren θ , sier vi at x_n er en konsistent estimator av θ .

Begrepet konvergens i sannsynlighet kan lett utvides til å omfatte stokastiske vektorer eller matriser. Vi sier at den stokastiske vektoren \underline{x}_n konvergerer i sannsynlighet mot \underline{x} dersom hvert element i \underline{x}_n konvergerer i sannsynlighet mot det tilsvarende element i \underline{x} .

A.3.5. Konvergens i fordeling

La x_1, x_2, \dots være en følge av stokastiske variable og la x være en stokastisk variabel. La $F_n(t)$ og $F(t)$ være de kumulative fordelingsfunksjonene til x_n og x , henholdsvis, dvs. $F_n(t) = P(x_n \leq t)$, $F(t) = P(x \leq t)$. Vi sier at x_n konvergerer mot x i fordeling (lov) dersom

$$(A.21) \quad \lim_{n \rightarrow \infty} F_n(t) = F(t)$$

for ethvert punkt t der $F(t)$ er kontinuerlig.

Vi skriver

$$(A.22) \quad x_n \xrightarrow{L} x.$$

Dersom x_n er definert som det aritmetiske gjennomsnittet av n uavhengige, identisk fordelte stokastiske variable med forventning μ og varians $\sigma^2 < \infty$, kan en vise at

$$(A.23) \quad \sqrt{n}(x_n - \mu)/\sigma \xrightarrow{L} N(0,1).$$

(Vi skriver, for korthets skyld, " $x_n \xrightarrow{L} N(0,1)$ " og mener " $x_n \xrightarrow{L} x$, der $x \sim N(0,1)$ ".) Vi sier da at x_n er asymptotisk normalfordelt med forventning μ og varians σ^2/n . Dette resultatet er kjent som sentralgrenseteoremet.

Begrepet konvergens i fordeling lar seg generalisere til å gjelde stokastiske vektorer og matriser med en simultan fordeling. Vi sier at x_n konvergerer i fordeling (lov) mot den stokastiske vektoren x dersom (A.21) gjelder med \underline{t} innsatt for t , der nå $F_n(\underline{t})$ og $F(\underline{t})$ står for de simultane kumulative fordelingsfunksjonene til x_n og x , henholdsvis.

Sentralgrenseteoremet kan likeledes generaliseres til å omfatte stokastiske vektorer. La x_n være (det aritmetiske) gjennomsnittet av n uavhengige, identisk fordelte stokastiske vektorer med forventning $\underline{\mu}$ og kovariansmatrise Σ . Da gjelder

$$(A.24) \quad \sqrt{n}(x_n - \underline{\mu}) \xrightarrow{L} N(0, \Sigma).$$

Bevis: se Dhrymes (1970), s. 107-108.

Det er mulig å vise at

$$(A.25) \quad x_n \xrightarrow{P} x \Rightarrow x_n \xrightarrow{L} x,$$

altså at konvergens i sannsynlighet impliserer konvergens i fordeling. (Bevis: se Rao (1973), s. 122.)

Det motsatte gjelder ikke generelt, men kun når x er en konstant. Mer generelt kan en vise

Setning A.1. La $x_n \xrightarrow{L} x$ og la g være en funksjon av x_n som med sannsynlighet 1 er kontinuerlig i punktet x . Da gjelder generelt

$$(A.26) \quad g(x_n) \xrightarrow{L} g(x)$$

og, dersom x er ikke-stokastisk,

$$(A.27) \quad g(x_n) \xrightarrow{P} g(x).$$

Bevis: se Billingsley (1968), s. 30-31.

Det følger spesielt av setning A.1 (A.27) at dersom $\hat{\Sigma}$ er en konsistent estimator av matrisen Σ , så er $\hat{\Sigma}^{-1}$ en konsistent estimator av Σ^{-1} , forutsatt at Σ^{-1} eksisterer.

Følgende resultat er kjent som Slutsky's teorem:

Setning A.2. La $x_n \xrightarrow{L} x$ og $y_n \xrightarrow{P} \gamma$ der γ er en konstant. Da gjelder

$$(A.28) \quad x_n + y_n \xrightarrow{L} x + \gamma$$

$$(A.29) \quad y_n \cdot x_n \xrightarrow{L} \gamma x$$

$$(A.30) \quad x_n/y_n \xrightarrow{L} x/\gamma \text{ hvis } \gamma \neq 0.$$

Bevis: se Rao (1973), s.122.

Setning A.2 kan brukes til å bevise

Setning A.3. La a_1, a_2, \dots være en ikke-stokastisk tallfølge slik at a_n går mot uendelig, la μ være en fast konstant, og anta

$$(A.31) \quad a_n(x_n - \mu) \xrightarrow{L} y.$$

La videre g være en deriverbar funksjon med kontinuerlig derivert g' i μ . Da gjelder

$$(A.32) \quad a_n[g(x_n) - g(\mu)] \xrightarrow{L} g'(\mu)y.$$

Bevis: se Bickel og Doksum (1977), s. 461-462.

Som et spesialtilfelle av setning A.3 kan vi sette opp

Setning A.4. La

$$(A.33) \quad \sqrt{n}(x_n - \mu) \xrightarrow{L} N(0, \sigma^2)$$

og la g være en slik funksjon som nevnt i setning A.3. Da gjelder

$$(A.34) \quad \sqrt{n}(g(x_n) - g(\mu)) \xrightarrow{L} N(0, [g'(\mu)]^2 \sigma^2)$$

Setning A.4 kan generaliseres til å gjelde multinormalfordelte vektorer:

Setning A.5. La

$$(A.35) \quad \sqrt{n}(x_n - \mu) \xrightarrow{L} N(0, \Sigma)$$

og la $g = (g_1, g_2, \dots, g_r)'$ være en vektor av funksjoner av x_n med kontinuerlige partielt deriverte i μ . La $\partial g / \partial \mu$ betegne matrisen hvor element (i, j) er den partielt deriverte av g_i m.h.p. j -te element i x_n , utregnet i punktet $x_n = \mu$. Da gjelder

$$(A.36) \quad \sqrt{n}(g(x_n) - g(\mu)) \xrightarrow{L} N(0, (\partial g / \partial \mu) \Sigma (\partial g / \partial \mu)').$$

Bevis: se Rao (1973), s.387.

En kan legge merke til at dette resultatet som et spesialtilfelle omfatter funksjoner av formen $g(x) = Ax$, der A er en konstant matrise. Her får vi $(\partial g / \partial \mu) = A$ og følgelig

$$(A.37) \quad \sqrt{n}A(x_n - \mu) \xrightarrow{L} N(0, A \Sigma A').$$

Vi kan nå vise

Setning A.6. La x_n være en r -dimensjonal stokastisk vektor med asymptotisk fordeling gitt ved

$$(A.38) \quad \sqrt{n}(x_n - \mu) \xrightarrow{L} N(0, \Sigma),$$

der Σ er ikke-singulær. La $\hat{\Sigma}$ være en konsistent estimator av Σ , altså

$$(A.39) \quad \text{plim } \hat{\Sigma} = \Sigma.$$

La videre K være en $(v \times r)$ ikke-stokastisk matrise med rang v . Da gjelder

$$(A.40) \quad n(x_n - \mu)' K' (K \hat{\Sigma} K')^{-1} K (x_n - \mu) \xrightarrow{L} \chi^2(v).$$

Bevis¹⁾: Ifølge setning A.5 (A.37) har vi

$$(A.41) \quad \sqrt{n} K(x_n - \mu) \xrightarrow{L} N(0, K \Sigma K').$$

Matrisen $K \Sigma K'$ (og $K \hat{\Sigma} K'$) har $v(v+1)/2$ distinkte elementer. La $v(K \Sigma K')$ (henh. $v_n(K \hat{\Sigma} K')$) være vektoren av disse elementene, og betrakt $y_n = [(\sqrt{n} K(x_n - \mu))' \quad v_n'(K \hat{\Sigma} K')]'$ som en vektor av $v + v(v+1)/2$ stokastiske variable. For denne vektoren gjelder da

$$(A.42) \quad y_n \xrightarrow{L} y \sim \begin{pmatrix} N(0, K \Sigma K') \\ v(K \Sigma K') \end{pmatrix},$$

der $v_n(K \hat{\Sigma} K') \xrightarrow{L} v(K \Sigma K')$ følger av (A.25). Definer funksjonen

$$(A.43) \quad g(y_n) = n(x_n - \mu)' K' (K \hat{\Sigma} K')^{-1} K (x_n - \mu).$$

Funksjonen g er kontinuerlig i punktet y siden K har full rang, slik at $K \Sigma K'$ er ikke-singulær. Vi kan derfor anvende setning A.1 (A.26), som, sammenholdt med (A.16), gir

$$(A.44) \quad g(y_n) \xrightarrow{L} \chi^2(v) \quad \text{q.e.d.}$$

Poenget med dette resultatet er følgende. Kovariansmatrisen Σ i den asymptotiske (normal-) fordelingen er i alminnelighet ukjent. Dersom imidlertid $\hat{\Sigma}$ er en konsistent estimator av Σ , er dette tilstrekkelig til at kvadratsummen (A.40) er asymptotisk χ^2 -kvadratfordelt. I avsnitt 4 og 5 er dette nyttet til å konstruere asymptotiske tester av generelle lineære hypoteser om parametrene i den underliggende sannsynlighetsfordelingen til utvalgsdataene.

A.3.6. Informasjonsmatrisen

La x være en stokastisk vektor med simultan tetthet $f(x, \theta)$. θ er en (skalar) parameter. Informasjonstallet til denne fordelingen er definert som

$$(A.45) \quad \psi(\theta) = -E \left[\frac{\partial^2 \log f(x, \theta)}{\partial \theta^2} \right].$$

$\psi(\theta)$ er et ikke-negativt tall.

Informasjonstallet er et spesialtilfelle av et mer generelt begrep. Dersom tettheten til x avhenger av en vektor av parametre θ , kan vi definere informasjonsmatrisen

$$(A.46) \quad \Psi(\theta) = -E \left[\frac{\partial^2 \log f(x, \theta)}{\partial \theta \partial \theta'} \right]$$

De annenderiverte, m.h.p. parametrene θ , av logaritmen til tettheten danner en (symmetrisk) matrise. Vi tar forventningen av denne og snur fortegnet.

Informasjonsmatrisen er ikke-negativ definit. I det følgende vil vi forutsette at $\Psi(\theta)$ er positiv definit og følgelig ikke-singulær. Dette vil holde i nesten alle praktiske tilfelle.

1) Ideen til dette beviset stammer fra Jon Helgeland.

Dersom \underline{x} består av n subvektorer $\underline{x}_1, \dots, \underline{x}_n$, der hver \underline{x}_j er én observasjon fra en simultan tetthet $f_j(\underline{x}_j, \theta) = f_1(\underline{x}_1, \theta)$, slik at $\underline{x}_1, \dots, \underline{x}_n$ er et utvalg av n uavhengige tilfeldige observasjoner fra samme stokastiske fordeling, kan vi skrive

$$(A.47) \quad \psi(\theta) = n\psi_1(\theta)$$

der

$$(A.48) \quad \psi_1(\theta) = -E \left[\frac{\partial^2 \log f_1(\underline{x}_1, \theta)}{\partial \theta \partial \theta'} \right].$$

$\psi_1(\theta)$ kalles gjerne informasjonsinnholdet i én observasjon ("the information contained in one observation"). (A.47) - (A.48) holder fordi når observasjonene er uavhengige, er den simultane tettheten for hele utvalget gitt ved produktet av tetthetene for de enkelte observasjoner:

$$(A.49) \quad f(\underline{x}, \theta) = \prod_{j=1}^n f_1(\underline{x}_j, \theta) = [f_1(\underline{x}_1, \theta)]^n.$$

A.3.7. Cramér-Rao's ulikhet

La \underline{x} være en stokastisk vektor med simultan tetthet $f(\underline{x}, \theta)$, og la $a(\underline{x})$ være en (skalar) observator med endelig varians for alle θ . La $\alpha(\theta)$ betegne forventningen til $a(\underline{x})$, og anta at $\alpha(\theta)$ er deriverbar m.h.p. θ med derivert $\alpha'(\theta)$. Under visse (nokså generelle) "regularitetsbetingelser" kan en vise Cramér-Rao's ulikhet:

$$(A.50) \quad \text{var}_{\theta}(a(\underline{x})) \geq \frac{[\alpha'(\theta)]^2}{\psi(\theta)}.$$

Bevis: se Bickel og Doksum (1977), s. 127-129.

Dersom $a(\underline{x})$ er forventningsrett for θ , altså $\alpha(\theta) = \theta$ for alle verdier av θ , har vi, som spesialtilfelle av (A.50),

$$(A.51) \quad \text{var}_{\theta}(a(\underline{x})) \geq 1/\psi(\theta).$$

Tallet $1/\psi(\theta)$ kalles Cramér-Rao's nedre grense og angir den laveste varians noen forventningsrett estimator kan ha. En forventningsrett estimator som når denne grensen er derfor optimal innenfor klassen av forventningsrette estimatore. Det er imidlertid ikke alltid mulig å finne en slik estimator.

Nå er det ikke gitt at vi ønsker å begrense oppmerksomheten til klassen av forventningsrette estimatore. Dersom vi sammenlikner estimatore som ikke alle er forventningsrette, vil vi gjerne bruke bruttovariansen ("mean square error"), og ikke variansen, som optimalitetskriterium. Bruttovariansen til en estimator $a(\underline{x})$ av θ er gitt ved

$$(A.52) \quad \begin{aligned} \text{bvar}_{\theta}(a(\underline{x})) &= E_{\theta} [a(\underline{x}) - \theta]^2 \\ &= \text{var}_{\theta}(a(\underline{x})) + s_{\theta}^2 \end{aligned}$$

der

$$(A.53) \quad s_{\theta} = \alpha(\theta) - \theta.$$

s_{θ} er skjevheten av $a(\underline{x})$ m.h.p. θ .

Den nedre grense for bruttovariansen til $a(\underline{x})$ er gitt ved

$$(A.54) \quad \text{bvar}_{\theta}(a(\underline{x})) \geq \frac{[\alpha'(\theta)]^2}{\psi(\theta)} + s_{\theta}^2.$$

Et resultat tilsvarende (A.50) gjelder også i flerparametertilfellet. La $\underline{a}(\underline{x})$ være en $(s \times 1)$ vektor av funksjonelt uavhengige observatorer, la $\underline{\theta}$ være en $(m \times 1)$ vektor av funksjonelt uavhengige parametre, og definer

$$(A.55) \quad \underline{\alpha}(\underline{\theta}) = E_{\underline{\theta}}[\underline{a}(\underline{x})]$$

$$(A.56) \quad \Sigma_{\underline{a}} = \text{Cov}_{\underline{\theta}}(\underline{a}(\underline{x}))$$

$$(A.57) \quad \Gamma = \frac{\partial \underline{\alpha}(\underline{\theta})}{\partial \underline{\theta}} = \left[\frac{\partial \alpha_i(\underline{\theta})}{\partial \theta_j} \right] \quad (i=1,2,\dots,s; j=1,2,\dots,m).$$

Γ er matrisen av partielt deriverte av $\underline{\alpha}(\underline{\theta})$ m.h.p. $\underline{\theta}$.

Vi kan da, analogt med (A.50), vise at matrisen

$$(A.58) \quad \Delta = \Sigma_{\underline{a}} - \Gamma[\Psi(\underline{\theta})]^{-1}\Gamma'$$

er ikke-negativ definit, dvs. at $\Sigma_{\underline{a}}$ er "minst like stor som" Cramér-Rao's nedre grense $\Gamma[\Psi(\underline{\theta})]^{-1}\Gamma'$ (se Dhrymes (1970), s. 125-126). Dette innebærer spesielt at ethvert diagonalelement i $\Sigma_{\underline{a}}$ er minst like stort som det tilsvarende element i $\Gamma[\Psi(\underline{\theta})]^{-1}\Gamma'$, og at enhver lineærkombinasjon $\underline{c}'\underline{a}(\underline{x})$ av $\underline{a}(\underline{x})$ har en nedre grense for sin varians gitt ved

$$(A.59) \quad \text{var}_{\underline{\theta}}(\underline{c}'\underline{a}(\underline{x})) \geq \underline{c}'\Gamma[\Psi(\underline{\theta})]^{-1}\Gamma'\underline{c}.$$

\underline{c} er her en vilkårlig $(s \times 1)$ vektor av konstanter.

Dersom $s = m$, og $\underline{a}(\underline{x})$ er en forventningsrett estimator av $\underline{\theta}$, dvs. $\Gamma = I$ for alle $\underline{\theta}$, reduseres

(A.58) til

$$(A.60) \quad \Delta = \Sigma_{\underline{a}} - [\Psi(\underline{\theta})]^{-1},$$

som er analog med (A.51). Den inverse av informasjonsmatrisen er Cramér-Rao's nedre grense for kovariansmatrisen til en forventningsrett estimator av parametervektoren $\underline{\theta}$.

A.3.8. Om asymptotisk optimalitet

I mange tilfelle er det umulig å etablere den eksakte fordelingen til en stokastisk observator. I slike tilfelle tyr en gjerne til asymptotisk teori, dvs. en undersøker hvilke fordelingsegenskaper som gjelder når utvalgsstørrelsen går mot uendelig. En antar videre at i "store" utvalg kan den asymptotiske fordelingen nyttes som en brukbar tilnærming til utvalgsfordelingen. Det er imidlertid som regel vanskelig å avgjøre hvor stort det "store" utvalget må være for at denne framgangsmåten kan forsvares.

Sentralgrenseteoremet, sammenholdt med setning A.5, innebærer at en meget stor del av de utvalgsobservatorer vi får å arbeide med i praksis, er asymptotisk (multi-)normalfordelt. Vi blir da ledet til å undersøke hvilken asymptotisk normalfordeling de enkelte observatorer har, og spesielt om denne fordelingen i noen forstand kan sies å være optimal.

Det viser seg at det eksisterer en ganske vid klasse av observatorer som, under visse nokså generelle "regularitetsbetingelser", er konsistente og asymptotisk (multi-)normalfordelte med kovariansmatrise som (asymptotisk) når Cramér-Rao's nedre grense Ψ^{-1} . Slike estimatorer kalles "best asymptotisk normalfordelte" ("best asymptotically normal" - BAN) eller asymptotisk effisiente. Dersom vi kan finne en konsistent estimator $\hat{\psi}$ for informasjonsmatrisen, så er $\hat{\psi}^{-1}$ derfor en konsistent estimator for kovariansmatrisen til BAN-estimatoren.

A.3.9. Sannsynlighetsmaksimeringsprinsippet (ML)

La x være en stokastisk vektor med simultan tetthet $f(x, \theta)$. θ er en parametervektor. Sannsynlighetsmaksimeringsprinsippet ("the maximum likelihood principle" - ML) for estimering av θ går ut på å maksimere $f(x, \theta)$ m.h.p. θ , gitt observasjonene x . Løsningen på dette maksimeringsproblemet, $\hat{\theta}$, kalles, hvis den finnes, sannsynlighetsmaksimeringsestimatoren (ML-estimatoren) av θ . Størrelsen $f(x, \theta)$ brukt på denne måten kalles "likelihood"-funksjonen, og skrives gjerne $L(\theta, x)$ for å markere at vi nå betrakter θ som den variable og x som den gitte vektor.

Når regularitetsbetingelsene er oppfylt, er ML-estimatoren asymptotisk effisient (BAN) (se f.eks. Dhrymes, s. 114-130).

Siden logaritme-funksjonen er monotont stigende, blir det det samme enten vi maksimerer $L(\theta, x)$ eller $\log L(\theta, x)$. Det siste alternativet medfører i alminnelighet at regnearbeidet blir atskillig lettere.

Anta at vi er interessert i å teste en hypotese H_0 om parametervektoren θ , mot et alternativ H_1 . La L_0 og L_1 være likelihood-funksjonen maksimert under henholdsvis H_0 og H_1 . Brøken L_0/L_1 kalles sannsynlighetskvoten ("the likelihood ratio"). Sannsynlighetskvotetesten består i å forkaste H_0 dersom L_0/L_1 er lav. Et meget kjent og anvendelig resultat er at, under svært generelle betingelser, så gjelder

$$(A.61) \quad -2 \log(L_0/L_1) \xrightarrow{L} \chi^2(v).$$

$-2 \log(L_0/L_1)$ kalles kji-kvadratobservatoren basert på sannsynlighetskvoten ("the log-likelihood ratio chi-square statistic"). Ifølge (A.61) får vi en asymptotisk sannsynlighetskvotetest på nivå α ved å forkaste H_0 når $-2 \log(L_0/L_1)$ overstiger $(1-\alpha)$ -fraktilen i kji-kvadratfordelingen med v frihetsgrader. v er her forskjellen i antall frie parametre under H_1 og H_0 , eller antall uavhengige relasjoner som trengs for å beskrive H_0 i forhold til H_1 .

LITTERATUR

LITERATURE

- Abramowitz, M. og I.A. Stegun (eds.) (1972): "Handbook of mathematical functions". U.S. Government Printing Office, Washington D.C.
- Anscombe, F.J. (1956): "On estimating binomial response relations". Biometrika, 43, s. 461-464.
- Amemiya, T. (1975): "Qualitative response models". Annals of Economic and Social Measurement, 4, s. 363-372.
- Amemiya, T. (1980): "The n^{-2} -order mean squared errors of the maximum likelihood and the minimum logit-chi-square estimator". Annals of Statistics, 8, s. 488-505.
- Amundsen, A. (1968): "Metoder i analysen av forbruksdata". Statistisk Sentralbyrå, Artikler nr. 6, 2. opplag.
- Amundsen, H.T. (1974): "Binary variable multiple regression". Scandinavian Journal of Statistics, 1, s. 59-70.
- Amundsen, H.T. (1976): "Binary regressions for a polytomeous regressand". Scandinavian Journal of Statistics, 3, s. 39-41.
- Berkson, J. (1944): "Application of the logistic function to bio-assay". Journal of the American Statistical Association, 41, s. 70-74.
- Berkson, J. (1951): "Why I prefer logits to probits". Biometrics, 7, s. 327-339.
- Berkson, J. (1953): "A statistically precise and relatively simple method of estimating the bio-assay with quantal response, based on the logistic function". Journal of the American Statistical Association, 48, s. 565-599.
- Berkson, J. (1955a): "Maximum likelihood and minimum χ^2 estimates of the logistic function". Journal of the American Statistical Association, 50, s. 130-162.
- Berkson, J. (1955b): "Estimate of the integrated normal curve by minimum normit chi-square with particular reference to bio-assay". Journal of the American Statistical Association, 50, s. 529-549.
- Berkson, J. (1956): "Estimation by least squares and by maximum likelihood". Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1, s. 1-11. University of California Press, Berkeley and Los Angeles.
- Berkson, J. (1957): "Tables for use in estimating the normal distribution function by normit analysis". Biometrika, 44, s. 411-435.
- Berkson, J. (1968): "Application of minimum logit χ^2 estimate to a problem of Grizzle with a notation on the problem of 'no interaction'". Biometrics, 24, s. 75-95.
- Berkson, J. (1980): "Minimum chi-square, not maximum likelihood!" Annals of Statistics, 8, s. 457-487.
- Bickel, P.J. og K.A. Doksum (1977): "Mathematical statistics: basic ideas and selected topics". Holden-Day, San Francisco.
- Billingsley, P. (1968): "Convergence of probability measures". John Wiley & Sons, New York.
- Bishop, Y.M.M., S.E. Fienberg og P.W. Holland (1975): "Discrete multivariate analysis: theory and practice". MIT Press, Cambridge, Mass.
- Blalock, H.M. (1972): "Social statistics". Second edition. McGraw Hill, Tokyo.
- Clausen, S.-E. (1977): "Analyse av multivariate frekvenstabeller. Innføring i L.A. Goodmans analyse-system, og brukerbeskrivelse for programmet ECTA". INAS, Arbeidsrapport nr. 25.
- Clickner, R.P. (1976): "On least squares estimation for categorical data". Communications in Statistics - Theory and Methods, A5, s. 1059-1064.
- Cox, D.R. (1970): "The analysis of binary data". Methuen, London.
- Cramér, H. (1946): "Mathematical methods of statistics". Princeton University Press, Princeton.
- Dhrymes, P.J. (1970): "Econometrics: statistical foundations and applications". Harper & Row, New York.

- Domencich, T.A. og D.McFadden (1975): "Urban travel demand. A behavioral analysis". North-Holland, Amsterdam.
- Fridstrøm, L. (1978): "Yrkesdeltaking 1977-2000. Revisjon av beregninger utført til Langtidsprogrammet 1978-81". Statistisk Sentralbyrå, Arbeidsnotater IO 78/28.
- Gart, J.J. og J.R. Zweifel (1967): "On the bias of various estimators of the logit and its variance with application to quantal bioassay". Biometrika, 54, s. 181-187.
- Goldfeld, S.M. og R.E. Quandt (1972): "Nonlinear methods in econometrics". North-Holland, Amsterdam.
- Goodman, L.A. (1970): "The multivariate analysis of qualitative data: interactions among multiple classifications". Journal of the American Statistical Association, 65, s. 226-256.
- Goodman, L.A. (1971): "The analysis of multidimensional contingency tables: stepwise procedures and direct estimation methods for building models for multiple classifications". Technometrics, 13, s. 33-61.
- Haldane, J.B.S. (1955): "The estimation and significance of the logarithm of a ratio of frequencies". Annals of Human Genetics, 20, s. 309-311.
- Haldorsen, T. (1976): "Forelesninger om avhengighetsmål i kontingenstabeller av universitetslektor Harald Goldstein". Statistisk Sentralbyrå, Arbeidsnotater IO 76/27.
- Haldorsen, T. (1977): "Om log-lineær analyse av flerveistabeller". Statistisk Sentralbyrå, Arbeidsnotater IO 77/46.
- Hellevik, O. (1971): "Forskningsmetode i sosiologi og statsvitenskap". Universitetsforlaget, Oslo.
- Hernæs, E., O. Ljones og O. Vannebo (1977): "Yrkesdeltaking 1975-2000. En dokumentasjon av beregninger utført i tilknytning til arbeidet med Langtidsprogrammet 1978-1981". Statistisk Sentralbyrå, Arbeidsnotater IO 77/13.
- Hitchcock, S.E. (1962): "A note on the estimation of the parameters of the logistic function, using the minimum logit χ^2 method". Biometrika, 49, s. 250-252.
- Lee, T.C., G.G. Judge og A. Zellner (1970): "Estimating the parameters of the Markov probability model from aggregate time series data". North-Holland, Amsterdam.
- Ljones, O. (1979): "Kvinnens yrkesdeltaking i Norge". Statistisk Sentralbyrå, Samfunnsøkonomiske studier nr. 39.
- McFadden, D. (1974): "Conditional logit analysis of qualitative choice behavior". S. 105-142 i P. Zarembka (ed.): "Frontiers in econometrics". Academic Press, New York.
- McFadden, D. (1976): "Quantal choice analysis: a survey". Annals of Economic and Social Measurement, 5, s. 363-390.
- McGillivray, R. (1970): "Estimating the linear probability function". Econometrica, 38, s. 775-776.
- Nerlove, M. og S.J. Press (1973): "Univariate and multivariate log-linear and logistic models". R-1306-EDA/NIH, The Rand Corporation, Santa Monica.
- Neyman, J. (1949): "Contribution to the theory of the χ^2 test". Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability, s. 239-273. University of California Press, Berkeley and Los Angeles.
- Nie, N.H., C.H. Hull, J.G. Jenkins, K. Steinbrenner og D.H. Bent (1975): "SPSS. Statistical Package for the Social Sciences". Second edition. McGraw Hill, New York.
- Page, E. (1977): "Approximations to the cumulative normal function and its inverse for use on a pocket calculator". Applied Statistics, 26, s. 75-76.
- Rao, C.R. (1973): "Linear statistical inference and its applications". Second edition. John Wiley & Sons, New York.
- Scheffé, H. (1959): "The analysis of variance". John Wiley & Sons, New York.

- Schmeiser, B.W. (1979): "Approximations to the inverse cumulative normal function for use on hand calculators". Applied Statistics, 28, s. 175-176.
- Statistisk Sentralbyrå (1978): "Arbeidsmarkedsstatistikk 1977". NOS A 958.
- Stevens, S.S. (1946): "On the theory of scales of measurement". Science 103, s. 677-680.
- Theil, H. (1969): "A multinomial extension of the linear logit model". International Economic Review, 10, s. 251-259.
- Theil, H. (1971): "Principles of econometrics". John Wiley & Sons, New York.
- Zellner, A. and T.H. Lee (1965): "Joint estimation of relationships involving discrete random variables". Econometrica, 33, s. 382-394.

Trykt 1980

- Nr. 80/1 Svein Longva, Lorents Lorentsen and Øystein Olsen: Energy in a Multi-Sectoral Growth Model Energi i en flersektors vekstmodell Sidetall 22 Pris kr 9,00 ISBN 82-537-1082-8
- 80/2 Viggo Jean-Hansen: Totalregnskap for fiske- og fangstnæringen 1975 - 1978 Sidetall 33 Pris kr 9,00 ISBN 82-537-1080-1
- 80/3 Erik Biørn og Hans Erik Fosby: Kvartalsserier for brukerpriser på realkapital i norske produksjonssektorer Sidetall 60 Pris kr 11,00 ISBN 82-537-1087-9
- 80/4 Erik Biørn and Eilev S. Jansen: Consumer Demand in Norwegian Households 1973 - 1977 A Data Base for Micro-Econometrics Sidetall 130 Pris kr 13,00 ISBN 82-537-1086-0
- 80/5 Ole K. Hovland: Skattemodellen LOTTE Testing av framskrivingsmetoder Sidetall 30 Pris kr 9,00 ISBN 82-537-1088-7
- 80/6 Fylkesvise elektrisitetsprognoser for 1985 og 1990 En metodestudie Sidetall 56 Pris kr 11,00 ISBN 82-537-1091-7
- 80/7 Analyse av utviklingen i elektrisitetsforbruket 1978 og første halvår 1979 Sidetall 22 Pris kr 7,00 ISBN 82-537-1129-8
- 80/8 Øyvind Lone: Hovedklassifiseringa i arealregnskapet Sidetall 50 Pris kr 9,00 ISBN 82-537-1104-2
- 80/9 Tor Bjerkedal: Yrke og fødsel En undersøkelse over betydningen av kvinners yrkesaktivitet for opptreden av fosterskader Occupation and Outcome of Pregnancy Sidetall 93 Pris kr 11,00 ISBN 82-537-1111-5
- 80/10 Statistikk fra det økonomiske og medisinske informasjonssystem Alminnelige somatiske sykehus 1978 Sidetall 65 Pris kr 11,00 ISBN 82-537-1119-0
- 80/11 John Dagsvik: A Dynamic Model for Qualitative Choice Behaviour Implications for the Analysis of Labour Force Participation when the Total Supply of Labour is latent En dynamisk teori for kvalitativ valghandling Implikasjoner for analyse av yrkesdeltaking når det totale tilbud av arbeid er latent Sidetall 25 Pris kr 9,00 ISBN 82-537-1152-2
- 80/12 Torgeir Melien: Ressursregnskap for jern Sidetall 56 Pris kr 9,00 ISBN 82-537-1138-7
- 80/13 Eystein Glattre og Ellen Blix: En vurdering av dødsårsaksstatistikken Feil på dødsmeldingene Evaluation of the Cause-of-Death-Statistics Sidetall 64 Pris kr 11,00 ISBN 82-537-1136-0
- 80/14 Petter Frenger: Import Share Functions in Input - Output Analysis Importandelsfunksjoner i kryssløpsmodeller Sidetall 41 Pris kr 9,00 ISBN 82-537-1143-3
- 80/15 Den statistiske behandlingen av oljevirkosomheten Sidetall 56 Pris kr 11,00 ISBN 82-537-1150-6
- 80/16 Ådne Cappelen, Eva Ivås og Paal Sand: MODIS IV Detaljerte virkningstabeller for 1978 Sidetall 261 Pris kr 15,00 ISBN 82-537-1142-5
- 80/18 Susan Lingsom: Dagbøker med og uten faste tidsintervaller: En sammenlikning basert på prøveundersøkelse om tidsnytting 1979 Open and Fixed Interval Time Diaries: A Comparison Based on a Pilot Study on Time Use 1979 Sidetall 31 Pris kr 9,00 ISBN 82-537-1158-1
- 80/19 Sigurd Høst og Trygve Solheim: Radio- og fjernsynsundersøkelsen januar - februar 1980 Sidetall 101 Pris kr 13,00 ISBN 82-537-1155-7
- 80/20 Skatter og overføringer til private Historisk oversikt over satser mv. Årene 1969 - 1980 Sidetall 53 Pris kr 11,00 ISBN 82-537-1151-4
- 80/21 Olav Bjerkholt og Øystein Olsen: Optimal kapasitet og fastkraftpotensial i et vannkraftsystem Sidetall 36 Pris kr 9,00 ISBN 82-537-1154-9
- 80/22 Rolf Aaberge: Eksakte metoder for analyse av to-vegstabellar Sidetall 80 Pris kr 11,00 ISBN 82-537-1161-1

Utkommet i serien Rapporter fra Statistisk Sentralbyrå (RAPP) (forts.) - ISSN 0332-8422

Trykt 1980 (forts.)

- Nr. 80/23 P. Frenger, E. S. Jansen and M. Reymert: Tariffs in a World Trade Model An Analysis of Changing Competitiveness due to Tariff Reductions in the 1960's and 1970's Sidetall 47
Pris kr 9,00 ISBN 82-537-1163-8
- 80/24 Jan Mønnesland: Bestandsuavhengige giftermålsrater Sidetall 50 Pris kr 9,00
ISBN 82-537-1167-0
- 80/26 Lasse Fridstrøm: Lineære og log-lineære modeller for kvalitative avhengige variable
Linear and Log-Linear Qualitative Response Models Sidetall 122 Pris kr 13,00
ISBN 82-537-1184-0
- 80/27 Aktuelle skattetal 1980 Current Tax Data Sidetall 43 Pris kr 9,00 ISBN 82-537-1194-8

Pris kr 13,00

Publikasjonen utgis i kommisjon hos H. Aschehoug & Co. og
Universitetsforlaget. Oslo, og er til salgs hos alle bokhandlere.

ISBN 82-537-1184-0
ISSN 0332-8422