

Sven Berg

Några synpunkter på ekologisk inferens och kommentarer till en bok

Jag har blivit ombedd att granska boken *Modern Political Ecological Analysis* redigerad av Sten Berglund och Sören Risbjerg Thomsen. Eftersom jag är statistiker till professionen och inte statsvetare kommer jag att i huvudsak hålla mig till statistiska problem som aktualiseras i samband med ekologisk analys av väljarbeteende. Det är naturligt att jag i mina kommentarer ägnar speciell uppmärksamhet åt uppsats nr 2 i boken, valideringsuppsatsen, där man gör ett försök till utvärdering av den av Thomsen introducerade metodiken för ekologisk inferens. Som en nödvändig bakgrund till mina synpunkter ger jag i första avsnittet här en kortfattad översikt av området ekologisk inferens speciellt med hänsyn till undersökningar av väljarbeteende baserade på aggregerade data. Avsnittet 2 ägnas huvudsakligen åt en närmare granskning av den i boken redovisade valideringen. Några avslutande kommentarer rundar av.

Allmänna synpunkter

Betrakta följande två tabeller med artificiella data.

Tabell 1. Kontingenstabeller som saknar cellfrekvenser

Kategori	Utfall/frekvenser		Kategori	Utfall/frekvenser		
	A	B		A	B	
1		12	1		10	
2		8	2		10	
	10	10		9	11	20

Vi är intresserade av sannolikheterna (benägenheterna) för A-utfall givet kategori, Sannolikhet (A-utfall/kategori 1) och Sannolikhet (A-utfall/Kategori 2). Tabellerna utgör då ett exempel på *ofullständiga data*: cellvärdena i tabellerna saknas. Finns det då någon statistisk metod som möjliggör att man kan skatta cellvärdena, dvs. rekonstruera de fullständiga tabellerna? För att fylla de tomma cellerna behöver vi en statistisk imputeringsprocedur. Givetvis krävs någon form av antagande om hur data i tabellerna genererats. Under antagande att samma binomialfördelningar gäller radvis i tabellerna är det möjligt att utnyttja statistisk standardmetodik. »Maximum likelihood«-metoden¹ tillämpad direkt leder emellertid redan i detta enkla exempel till prohibitiva beräkningar. Den i sin enkelhet attraktiva EM-metoden (E = »expectation«, M = »maximization«) visar

sig konvergera mycket långsamt till lösningar. Kvasi-maximum-likelihood-metodik (se till exempel McCulagh och Nelder, 1989: 175) ger följande tabell med skattade betingade sannolikheter:

Tabell 2. Uppskattade eller rekonstruerade cellfrekvenser

Kategori	Utfall		
	A	B	
1	0.7	0.3	1.0
2	0.2	0.8	1.0

Ytterligare en möjlighet utgörs av »maximum entropy«-teknik (se Kaipur, 1990).² Startpunkten härvid är tabellernas fixa marginalvärden och man frågar sig vilka cellvärden som är minst »unbiased« (= representerar maximal entropi) med hänsyn till den kunskap, dvs. de restriktioner vi har i form av givna marginaler.

Vi kan se exemplet ovan som ett förenklat exempel på analys av aggregerade data, eller ekologisk analys med syfte att beskriva individuellt beteende. Individuella data som skulle gett cellfrekvenser och underlag för skattning direkt saknas. Vad inträffar nu i praktiken, till exempel vid ekologisk valforskning, där (1,2) kan representera socialgrupper och (A,B) politiska partier?³

I. Vi har att göra med *större urval*, eftersom våra ekologiska enheter (kommuner, valkretsar etc.) som regel omfattar tusentals individer, samt dessutom större och fler tabeller, dvs. *fler parametrar* att skatta: Vid danska val 12 olika partier + ytterligare ett par kategorier. I Sverige kanske 7 olika partier och mer än 270 kommuner, dvs. i runda tal 14 000 parametervärden att skatta!

II. *Heterogenitet* råder mellan tabeller, dvs. vi kan inte realistiskt förutsätta att samma politiska och sociala mekanismer driver fram data i de olika tabellerna. Lokala politiska kulturer och lokala valvindar kan som bekant skilja sig avsevärt från ett nationellt genomsnitt.

Punkten I medför att vi inte kan tillämpa statistiska standardmetoder och vanlig programvara, till exempel maximum likelihood-teknik, ens med tillgång till de kraftfullaste datorer. Approximativa tekniker måste tillgripas. Rent beräkningsmässiga problem träder i förgrunden. Under punkten II kan man naturligtvis tänka sig att försöka gruppera de ekologiska enheterna i homogena regioner, vilket emellertid förutsätter någon kunskap om den bakomliggande grupperingsmekanismen. Ur statistisk synpunkt ter det sig emellertid naturligare att försöka modellera heterogeniteten i väljarbeteendet.

I sociologisk forskning utnyttjades länge korrelation mellan rad- och kolumnvariabel beräknad över ekologiska enheter (= tabeller) för att bestämma ett genomsnittligt cellvärde, dvs. genomsnittlig betingad frekvens. Robinson visade i sin berömda och ofta citerade uppsats från 1950 vilka felslut som härvid kan uppstå. Hans paradexempel var den på rastillhörighet betingade frekvensen läskunnighet beräknad på censusdata (grupperade data), vilken visade sig avvika

kraftigt från motsvarande, korrekta värde beräknat på individdata. Problematiken som Robinson så effektivt demonstrerade har kommit att kallas »det ekologiska felslutet« (»the ecological fallacy«).

Vid forskning om väljarbeteende föreligger god tillgång på aggregerade data för administrativa och geografiska enheter, ofta av synnerligen god kvalitet, medan valhemligheten ställer hinder i vägen för ett direkt iakttagande av individuellt beteende. Intervjuundersökningar (surveys) ger visserligen individuella data, men datas kvalitet kan ibland ifrågasättas. Över ett längre tidsperspektiv är man naturligtvis helt hänvisad till aggregerade data. Det ekologiska problemet är således i högsta grad relevant i detta sammanhang!

Inom ekologisk valforskning har vi huvudsakligen att göra med tre tabelltyper representerande ofullständiga data av typ som ovan:

- Röstflödesmatriser för två val (som regel två på varandra följande) representerande kolumn respektive rad i tabellerna.
- Klassröstande: varvid en indelningsgrund utgörs av samhällsklass eller yrkeskategori och den andra partitillhörighet.
- Röstsplittring: med till exempel rad svarande mot kommunalval och kolumn mot riksdagsval.

De metoder vi finner i litteraturen för att komma till rätta med »det ekologiska problemet« som det här formuleras innehåller typiskt två komponenter:

- *Beräkningsmetod* (algoritm) för att rekonstruera celler i flödes- (bytes-)matrisen (matriserna).
- Någon form av *statistisk modell* (eller argumentation) åberopas för att motivera beräkningsmetoden.

Till den statistiska grundmodellen fogas som regel ett antal förutsättningar/villkor och förenklande antaganden, till exempel homogena (med konstanta väntevärden) bytes-benägenheter tvärs över ekologiska enheter. Detta för att man överhuvud skall komma fram till en beräkningsmetod som är möjlig att implementera i praktiken.

Vi finner i litteraturen ett antal metoder eller ansatser, i huvudsak nämligen:

- *regressionsansats* à la Goodman med modifikationer av typ »krympning« (»shrinkage«) eller restriktioner (som i Lewins (1972) arbete);
- *maximum entropiansats*, vilken vi förknippar med Johnstons (1985) arbeten;
- *multinomialansats*, som speciellt utvecklats av Brown och Payne (1986) och vidareutvecklats av italienarna Forcina och Marchetti (1990);
- *ekologisk korrelationsansats*, utarbetad av Thomsen i avhandlingen (1987) och andra arbeten och som står i centrum i den aktuella boken.

Till dessa fyra ansatser sällar sig nu möjligen Kohlsches (1990), komponentmetod vars detaljer jag tyvärr ännu inte hunnit sätta mig in i.

Det kan vara på sin plats att kortfattat kommentera dessa ansatser. I den aktuella boken ges ytterst sparsam, eller inte helt korrekt information om dessa metoder.

Att Goodmans regressionsmetod vid tillämpning på den typ av bytes- eller flödesmatriser vi här har att göra med leder till svårigheter är välbekant. Även vid modifiering av typ krympning av estimat uppstår ofta svårigheter i form av inadmissibla eller i vart fall svårtolkade resultat.⁴ Metodens popularitet får kanske tillskrivas dess enkelhet?

Maximum entropi representerar en generell metodik med betydligt större räckvidd än vad som framgår av litteraturen på det aktuella området (till exempel Johnstons och hans medarbetares arbeten). Som Johnston utnyttjar metodiken är den att betrakta som »kvasiekologisk«, eftersom man utnyttjar surveydata för att skatta nationella flöden, vilka kommer att utgöra sido-restriktioner vid beräkningar. Beräkningarna blir vid denna ansats ingenting annat än iterativa, proportionella anpassningar (Deming-Stephans metod). Vi får dessutom en överensstämmelse mellan maximum likelihood-skattning (under antagande om försumbart samspel av visst slag) och maximum entropi-metoden, vilket emellertid inte nödvändigtvis garanterar tillförlitliga resultat!⁵ I princip kan man emellertid tänka sig metodiken tillämpad på ett alternativt sätt vid ekologisk valforskning utan utnyttjande av surveydata och mera i linje med vad Herniter föreslagit vid analys av marknadsandelar (se utmärkt redogörelse i Kaipur (1989)).

Den av Brown och Payne (1986) introducerade aggregerade kumulativa multinomialmodellen (ACM, med deras förkortning) representerar en *integrerad statistisk modell*. Vi har ett fördelningsantagande som både förklarar utfall i enskild tabell: multinomialt radvis som »kumuleras« (dvs. summeras) till kända kolumnsummor och variation mellan tabeller via sannolikhetsvärden i multinomialfördelningarna genererad av Dirichletfördelningar, vars parametrar kan göras beroende av bakgrundsvariabler via logistisk regression. I princip skall ACM-modellen i sin egenskap av integrerad modell medge anpassningstest (»goodness-of-fit«), hypotes-prövning av närvaron av vissa effekter, konfidensintervall etc., vilket naturligtvis är en utomordentligt stor fördel. Den sofistikerade AMC-modellen⁶ är emellertid beräkningstekniskt ytterst krävande och en del utvecklingsarbete på programsidan torde återstå. Vi kanske kan hoppas att så småningom få tillgång till lätt implementerbara program för PC-bruk?

Vid Thomsens ekologiska logitmetod åberopas en statistisk grundmodell (giltig inom en »politiskt homogen« region) med latent variabelstruktur: givet individens/väljarens position på den latent variabeln är de två indelningsgrunderna statistiskt oberoende. Man slipper härigenom ifrån att explicit modellera beroendet mellan de två indelningsgrunderna, ett beroende som ju som regel är utomordentligt starkt (till exempel det mellan två på varandra följande val). Det beroende vi observerar hos de manifesta (observerade) variablerna åstadkoms/förklaras via den latent variabelns variation över ekologiska enheter. Två ytterligare, förenklande antaganden om isomorfism respektive variansförhållande individuellt/gruppvis ger grundval för en praktiskt användbar beräkningsmetod. Efter val av ett så kallat referensparti dikotomiseras tabellerna successivt till ett antal

2×2-tabeller. Vid tillämpning av metoden har valet av referensparti visat sig ha stor betydelse. För varje sådan uppsättning 2×2-tabeller beräknas korrelationer mellan logits av marginalfrekvenser (alltså tetrakorisk korrelation, inte den direkta Pearsonska produktmoment-korrelationen som Robinson varnade för i sin berömda uppsats!). Korrelationerna transformeras därefter till skattningar av cellvärden. Beskrivningen av metodiken i Thomsens avhandling (1987) är inte alldeles lätt att följa.

Det förtjänar framhållas att beräkningsmetoden är sådan, att det aldrig kan uppträda icke admissibla skattningar i form av negativa sannolikheter eller sannolikheter större än ett (vilket tenderar att inträffa vid Goodmans ekologiska regression). Thomsen har förtjänstfullt utvecklat programvara för PC, som han välvilligt ställer till andra forskares förfogande. Thomsens metod har väckt stort intresse och på kort tid fått stor spridning.

Vilken av de uppräknade metoderna är nu den bästa? Vilken metod leder till de mest tillförlitliga uppskattningarna? Det är inte alldeles lätt att avgöra. Ibland är syftet med beräkningarna olika, vilket naturligtvis medför olika kriterier för valideringen. Johnstons variant av maximum entropi-metoden syftar till att ge skattningar av lokala bytesmatriser, vilka därefter utnyttjas för en modellering och analys av regional variation i bytesbenägenheter. För Thomsen är målsättningen däremot en tillförlitlig skattning av en nationell bytesmatris. Grundläggande antaganden och förutsättningar skiljer sig åt. Statistiska anpassningstest (goodness-of-fit) förutsätter att man håller sig till en väl specificerad, begränsad klass av fördelningar (modeller) och lämpar sig inte för utvärdering av till exempel Thomsens metodik, men väl för ACM-modellen. Syftet här är inte att jämföra ekologiska logitmetoden med andra metoder, utan att försöka avgöra om metoden överhuvud ger användbara resultat. Garanterat »sanna värden« på partibytesbenägenheter disponerar vi givetvis inte och tillgången på lämpliga valideringssampel från intervjuundersökningar är begränsad. Artificiella data genererade vid simulering är ett alternativ som erbjuder vissa möjligheter till validering.⁷

Validering av ekologiska logitmetoden

Jag övergår nu till att närmare kommentera de båda inledande uppsatserna i den aktuella boken.

Sten Berglund skrev en gång en alldeles briljant kritik i *Statsvetenskaplig Tidskrift* av Leif Lewins (1972) bok om den svenska väljarkåren. Berglunds kritik, som tål att läsas än i dag, tog sikte på den något ytliga behandlingen av det ekologiska problemet i Lewins bok. Jag hade väntat mig mer av Berglunds introducerande uppsats i den aktuella boken.

Jag går nu över till den viktiga uppsatsen 2 i boken, där man försöker utvärdera den av Thomsen introducerade tekniken för ekologisk inferens. Man utnyttjar härvid som valideringssampel tillgängliga urvalsundersökningar (surveydata) från Danmark, Sverige och Finland. Problemet med surveydata är att de innehåller två felkomponenter: ett samplingfel, vilket är beräkningsbart, och med stor sannolikhet en biaskomponent (systematiskt fel), som vi inte känner. Den statistiska metodik författarna använder vid valideringen utgörs av en indexberäkning

på avvikelser mellan imputerade cellfrekvenser och surveybaserade frekvenstal. Härtill kommer vad man kallar en varianskomponentmodell för bedömning av de erhållna indexvärdena. Jag skall här peka på några problem vid valideringen som är speciellt iögonenfallande.

Stort antal parametrar – små cellvärden – stor osäkerhet

Man väljer att beräkna ett index för överensstämmelse mellan ekologiska skattningar och surveyskattning som baserar sig på jämförelse av relativa flöden, varvid totalsumman över samtliga celler blir 100 procent (i praktiken < 100 procent, eftersom man behöver jämförbara kategorier). När man sedan utnyttjar de ekologiska skattningarna i efterföljande uppsatser är det övergångsmatrisen med bytesbenägenheterna (och även rekryteringen) som intresserar, dvs. man räknar om till radsumma = 100 procent (respektive kolumnsumma = 100). Hade det inte varit naturligare att göra valideringen med utgångspunkt från de skattningar som faktiskt används i den statsvetenskapliga analysen av de skattade tabellerna? Man hade då sluppit ett antal irriterande små tal.

Med relativa flöden får man tabeller med ett stort antal mycket små siffror, typ 0.3 procent och liknande, eftersom huvudparten av väljarna befinner sig i huvuddiagonalen (byter ej parti). I det danska fallet har man ca. 60 procent i huvuddiagonalen och resterande 40 procent skall fördelas på ett mycket stort antal celler. Med 12 partier deltagande i bägge de undersökta valen har vi 132 celler utanför huvuddiagonalen representerande olika partibyten! Det säger sig självt att resultatet måste bli en flödestabell med ett stort antal nollor eller mycket små tal. Så visar sig också vara fallet (se sid 18 ff.).

Jag vet inte om författarna gjort klart för sig vilken kraftig samplingvariation (osäkerhet) man måste räkna med i de surveyskattningar av flöde som används vid valideringen. Man ger i varje fall inte läsaren någon uppfattning härom. Tag till exempel de ca. 10 danska kommunister (symbol K) från valet 1971 som kommit med i Observaurvalet, varav ingen bytt parti till 1973. Den observerade partiloyaliteten är alltså 100 procent. Men om vi sänker lojaliteten till säg 85 procent (= en bytessannolikhet på 0.15) har vi fortfarande en 20 procentig sannolikhet att alla 10 stannar kvar i partiet. För de små delurval som uppstår i flerpartisystem av skandinavisk typ (speciellt Danmark) måste vi räkna med en rejäl felmarginal vid skattning av bytesbenägenheter för småpartierna!

Jag skulle gärna sett att man redovisat en enklare och mindre »ambitiös« utvärdering av Thomsens metod, som inte i så hög grad byggt på dessa yttest små, skattade frekvenser, men där man i gengäld skulle kunnat utnyttja etablerade metoder för statistisk inferens. Här följer ett förslag.

Ett grovt mått på lojaliteten/rörligheten i valmanskåren är relativa andelen som röstar med samma parti vid de två valtillfällena, dvs. i stort sett andelen som befinner sig i flödestabellens huvuddiagonal.⁸ Från Observatabellen på sid 19 har jag 58.0 procent som röstar med samma parti av ca 925 tillfrågade väljare (73.4 procent av 1260). Ett 95 procentigt konfidensintervall för andelen lojala väljare blir då: 54.8 – 61.2. Motsvarande skattning av lojaliteten beräknad med Thomsens metod (sid 18) för ekologisk skattning blir 62.5 procent. Som vi ser är det inte

troligt att skillnaden i uppskattad partiloyalitet kan tillskrivas samplingvariationen i Observaurvalet. (Jag följer Thomsen et al. och antar att de ekologiskt skattade värdena är utan slumpfel). Det föreligger uppenbarligen en systematisk skillnad. På motsvarande sätt finner man i både det finska och svenska fallet en högre skattad lojalitet med den ekologiska metoden.

Om detta förhållande – överskattning av partiloyaliteten – gäller generellt för Thomsens metod låter jag vara osagt. Syftet här har varit att visa på en enkel men ändå någorlunda korrekt statistisk metod att jämföra de olika sätten att skatta väljares bytesbenägenhet.

Varianskomponentmodellen för validering

Thomsen et al. utvecklar vad man kallar en varianskomponentmodell för jämförelse mellan surveybaserade flödesskattningar och ekologiskt skattade flöden. De senare antas vara utan slumpfel men (eventuellt) innehålla ett systematiskt fel. Surveyskattningarna å sin sida innehåller oundvikligen ett slumpfel (= urvalsfel) och därtill (förmodligen) en bias. Under ideala omständigheter – »an ideal survey« – kan man kanske bortse från biaskomponenten i surveyskattningarna. Detta är de grundantaganden som görs för modellen som används vid valideringen.

För danskt vidkommande disponerar man för valen 1971-73 två urvalsundersökningar och man kan alltså dels jämföra surveys inbördes, dels jämföra ekologiska resultat med surveyresultat. Om vi nu jämför surveys med varandra (»ideala surveys«), får vi räkna med att urvalsfelen adderas. (Mer precist: variansen för en skillnad beräknad på de två surveys är summan av två varianser). Jämför vi istället resultat från en survey med ekologiska resultat under antagande om »ideal survey« och inga slumpfel i de ekologiska resultaten, så får vi en avvikelse (ett indexmått), vars varians enbart består av en komponent: slumpfelet från surveyn. Om vi observerar en större diskrepans (= högre värde på avvikelseindex) mellan surveys inbördes än mellan ekologiska resultat och surveysiffror, så indikerar detta enligt författarna att den ekologiska metoden ger valida resultat.

Argumentationen är inte helt övertygande. Så här verkar författarna resonera: $\text{Varians}(X) > \text{Varians}(Y)$. Vi observerar ett värde större än ett annat. Det större värdet måste då vara ett X-värde! Varianser, dvs. statistisk variabilitet, får emellertid inte tolkas deterministiskt på det sätt som författarna gör. Vid en enda genomförd jämförelse ger inte en större varians garanti för att man skall observera en större avvikelse, speciellt inte om det rör sig om ungefärligen dubblad varians, dvs. ca. 40 procent högre spridning.

Om författarna konsulterat en statistiker hade man säkert kunnat få hjälp med att utveckla en metod för signifikansprövning av det avvikelseindex man använder sig av vid valideringen. Ovanan att handskas med statistiska grundbegrepp lyser igenom.

Ytterligare frågor

Det återstår ytterligare ett antal frågor och problem rörande Thomsens metod för ekologisk inferens som lämnas obesvarade i boken, men som är av stort intresse med tanke på metodens framtida användning vid ekologisk inferens.

En sådan fråga gäller val av homogena regioner inom vilka modellens grundantaganden antas gälla. Hur känslig (eller robust) är metoden härvid? Varför redovisar man inte resultat av skattningar för de sju regioner man indelat Danmark i? Vid en utvärdering av metodiken förfaller det rimligt att man prövar om man lyckats väl med denna inledande gruppering. Hur viktig är det att man finner fram till »politiskt homogena« regioner?

Till de obseverade frågorna hör också valet av referensparti vid beräkningarna av tetrakorisk korrelation. Erfarenheter vi gjort i Lund visar att valet är betydelsefullt. Metoden är inte robust i detta avseende.

Avslutande kommentarer

Sören Risbjerg Thomsen har gjort en aktningsvärd, kreativ insats genom att upptäcka användbarheten hos latent variabelmodeller vid ekologisk inferens. Han har härvid arbetat i en fin dansk tradition initierad av Rasch. Thomsens modell för ekologisk inferens är emellertid inte en sammanhållen, eller integrerad modell i samma mening som Brown och Paynes ACM-modell. Det är möjligt att man kan utveckla modellen vidare teoretiskt. Härför krävs samarbete med statistisk expertis.

Ekologisk inferens tilldrar sig allt större intresse. Med ett interdisciplinärt samarbete omfattande statsvetare, statistiker, geografer och kanske särskilt ekonometriker kan vi räkna med ytterligare framsteg på området och fördjupad kunskap. Berglund och Thomsen har lämnat ett viktigt bidrag med den bok de redigerat. Boken innehåller ett stort antal uppsatser som visar hur användbara aggregerade data kan vara.

Jag skulle avslutningsvis vilja peka på ett område, där man utvecklat ganska sofistikerade statistiska modeller, som skulle kunna finna tillämpning även inom valforskningen. Jag tänker på *marketing* och vad som där går under beteckningen »discrete choice models«, utvecklade av Ben Akiva och andra (se, till exempel, Ben-Akiva och Lerman, 1985).

Noter

1. Se till exempel King (1989) som ser likelihood-teorin som ett sätt att förenhetliga den politiska metodologin.
2. Även om merparten av tillämpningarna i Kaipurs bok är naturvetenskapliga innehåller den stort antal exempel på samhällsvetenskapliga problem, där maximum entropi metoden är tillämpbar, blant annat just ekologisk analys av partibyten.
3. Det är intressant att notera att McCullagh och Nelder i 2:a upplagan av sin bok (1989: 336 ff.) exemplifierar problematiken med skattning av sannolikheter från tabellmarginaler just med övergångar mellan partier vid val (»voter transitions«).
4. En intressant tillämpning av ekologisk regression kan noteras vid amerikanska domstolar i samband med mål rörande diskriminering av minoriteter. För att leda i bevis att en valkretsindelning faktiskt missgynnar en minoritet utnyttjas ekologisk regression (se Freedman et al., 1991).
5. För viss diskussion och ett blygsamt försök till validering av metoden se Berg (1988).
6. ACM-modellen är svårtillgänglig för icke-specialister. På sid 114 i den aktuella boken sägs modellen utgöra »..effort ... to employ better estimators..«, ett klart understatement.

7. Vid en workshop vid ECPR-mötet i mars 1991 presenterades resultat från sådana simuleringsförsök. Dessa resultat var naturligtvis inte tillgängliga för författarna till den aktuella boken.
8. Ett ytterligare grovt mått är naturligtvis den från engelsk litteratur välkända »swingen«.

Litteratur

- Ben-Akiva, M. och S. R. Lerman (1985). *Discrete Choice Analysis*, Cambridge: MIT Press.
- Berg, S. (1988). »Spatial influence on voter transitions in Swedish elections: an application of Johnston's maximum entropy method«, *Electoral Studies*, pp. 233-250.
- Brown, P. J. och C. D. Payne (1986). »Aggregate data, ecological regression, and voting transitions«, *Journal of the American Statistical Association*, 81, pp. 452-460.
- Forcina, A. och G. M. Marchetti (1991). *Effect of misspecification in Brown and Payne model*«, Paper presented at the ECPR meeting, Colchester.
- Freedman, D. A., S. Klein, J. Sacks, C. Smith och C. Everett (1991). *Ecological regression and voting rights*, Technical Report nr. 248, Department of Statistics, University of California, Berkeley.
- Johnston, R. J. (1985). *The Geography of English Politics*, London: Croom Helm.
- Kaipur, J. N. (1989). *Maximum Entropy Models in Science and Engineering*, Chichester: John Wiley & Sons.
- King, G. (1989). *Unifying Political Methodology*, Cambridge: Cambridge University Press.
- Kohlsche, A. J. (1990). *Forschungsbericht Nr. 1 zu Methoden der empirischen Wahlforschung*.
- Lewin, L. et al. (1972). *The Swedish Electorate 1887-1968*, Stockholm: Almqvist & Wiksell.
- McCullagh, P. och J. A. Nelder (1989). *Generalized Linear Models*, Chapman and Hall, 2:a upplagan.
- Thomsen, S. Risbjerg (1987). *Danish Elections 1920-79. A logit approach to ecological analysis and inference*, Århus: Forlaget Politica.