

Analyse af konjunkturbarometret for industrien

Carsten Stig Poulsen

Institut for Produktion, Aalborg Universitetscenter

SUMMARY: The article proposes a general methodology, viz. latent structure analysis, for analyzing the business cycle, using qualitative indicators of the IFO Institute-type. As a special case the latent class model is applied to Danish data and demonstrated to provide useful insights into the structural relation between the indicators and an effective summary of an otherwise complex data set.

Indledning

Som en del af konjunkturstatistikken udarbejder Danmarks Statistik det såkaldte »Konjunkturbarometer for industrien.« Undersøgelsen foretages kvartalsvist og bygger på konjunkturbedømmelser fra ledelserne af ca. 650 større industrivirksomheder, der tilsammen repræsenterer ca. 60% af beskæftigelsen i industrien. Offentliggørelse finder sted efter 3-4 uger i *Nyt fra Danmarks Statistik*, senere også i *Statistikservice* og *Statistiske Efterretninger*. Endelig indgår resultaterne i publikationen *Konjunkturoversigt*, som produceres 4 gange årligt.

Sammenlignet med andre økonomiske tidsserier er datamaterialet for »Konjunkturbarometeret« af flere grunde usædvanligt. *For det første* er der tale om en surveyundersøgelse, hvor de *samme respondenter* optræder i hver spørgerunde. Respondenterne, det vil her sige virksomhedsledelserne, siges at udgøre et *panel*. Almindeligvis vil stikprøveundersøgelser, således som de gennemføres af Danmarks Statistik i forbindelse med f.eks. forbrugsundersøgelserne, omfatte et nyt udvalg af respondenter fra gang til gang. Panelundersøgelser giver særlige muligheder for at belyse forskydninger over tiden. Mens gentagne stikprøver ved betragtning af den marginale fordeling på en række kriterier kun giver mulighed for vurdering af *nettobevægelser*, er det ved paneldata tillige muligt at analysere *bruttobevægelserne*. Man undersøger her den simultane fordeling af disse kriterier på to eller flere

Nærværende arbejde blev påbegyndt, mens forfatteren var seniorstipendiat ved Handelshøjskolen i Århus. Fuldmægtig Peter Dal, Danmarks Statistik, takkes for velvillig bistand i forbindelse med fremskaffelsen af de anvendte data.

Analyse af konjunkturbarometret for industrien

Carsten Stig Poulsen

Institut for Produktion, Aalborg Universitetscenter

SUMMARY: The article proposes a general methodology, viz. latent structure analysis, for analyzing the business cycle, using qualitative indicators of the IFO Institute-type. As a special case the latent class model is applied to Danish data and demonstrated to provide useful insights into the structural relation between the indicators and an effective summary of an otherwise complex data set.

Indledning

Som en del af konjunkturstatistikken udarbejder Danmarks Statistik det såkaldte »Konjunkturbarometer for industrien.« Undersøgelsen foretages kvartalsvist og bygger på konjunkturbedømmelser fra ledelserne af ca. 650 større industrivirksomheder, der tilsammen repræsenterer ca. 60% af beskæftigelsen i industrien. Offentliggørelse finder sted efter 3-4 uger i *Nyt fra Danmarks Statistik*, senere også i *Statistikservice* og *Statistiske Efterretninger*. Endelig indgår resultaterne i publikationen *Konjunkturoversigt*, som produceres 4 gange årligt.

Sammenlignet med andre økonomiske tidsserier er datamaterialet for »Konjunkturbarometeret« af flere grunde usædvanligt. *For det første* er der tale om en surveyundersøgelse, hvor de *samme respondenter* optræder i hver spørgerunde. Respondenterne, det vil her sige virksomhedsledelserne, siges at udgøre et *panel*. Almindeligvis vil stikprøveundersøgelser, således som de gennemføres af Danmarks Statistik i forbindelse med f.eks. forbrugsundersøgelserne, omfatte et nyt udvalg af respondenter fra gang til gang. Panelundersøgelser giver særlige muligheder for at belyse forskydninger over tiden. Mens gentagne stikprøver ved betragtning af den marginale fordeling på en række kriterier kun giver mulighed for vurdering af *nettobevægelser*, er det ved paneldata tillige muligt at analysere *bruttobevægelserne*. Man undersøger her den simultane fordeling af disse kriterier på to eller flere

Nærværende arbejde blev påbegyndt, mens forfatteren var seniorstipendiat ved Handelshøjskolen i Århus. Fuldmægtig Peter Dal, Danmarks Statistik, takkes for velvillig bistand i forbindelse med fremskaffelsen af de anvendte data.

Ved besvarelsen af spørgsmålene bedes der set bort fra normale sæsonbestemte ændringer.							
Udviklingsforløb	Faktisk forløb			Förventet forløb			
	i forhold til			i forhold til			
	Mindre	Omtrent uændret	Større	Mindre	Omtrent uændret	Større	
1. Produktion målt i væremængder	21	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
2. Beskæftigelse målt i antal ansatte ultimo kvartalene	24	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
3. Samlet ordreindgang målt i væremængder	27	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
4. Samlede løbende investeringer målt i værdi. Lagerændringer og køb af værdipapirer medregnes ikke	31	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
Vurdering af forholdene ultimo tællingskvartalet							
5. Samlet ordrebeholdning vurderet i væremængder. Virksomheder uden ordrebeholdning bedes i stedet vurdere styrken af efterspørgslen	41	<input type="checkbox"/> 1 - Større end normal	<input type="checkbox"/> 2 - Normal	<input type="checkbox"/> 3 - Mindre end normal			
6. Eksportordrebeholdning vurderet i væremængder. Eksporterende virksomheder uden eksportordrebeholdning bedes i stedet vurdere styrken af eksportefterspørgslen	43	<input type="checkbox"/> 1 - Større end normal	<input type="checkbox"/> 2 - Normal	<input type="checkbox"/> 3 - Mindre end normal	<input type="checkbox"/> 4 - Ingen eksport		
7. Færdigvarelagre vurderet i væremængder	45	<input type="checkbox"/> 1 - Større end normal	<input type="checkbox"/> 2 - Normal	<input type="checkbox"/> 3 - Mindre end normal	<input type="checkbox"/> 4 - Ingen færdigvarelagre		
8. Råvarelagre vurderet i væremængder	47	<input type="checkbox"/> 1 - Større end normal	<input type="checkbox"/> 2 - Normal	<input type="checkbox"/> 3 - Mindre end normal	<input type="checkbox"/> 4 - Ingen råvarelagre		
9. Produktionskapacitet	49	<input type="checkbox"/> 1 - Mere end tilstrækkelig	<input type="checkbox"/> 2 - Tilstrækkelig	<input type="checkbox"/> 3 - Ikke tilstrækkelig			
10. Blev planerne for fremtidige investeringer ændret i tællingskvartalet?	51	<input type="checkbox"/> 1 - Ja, udvidet	<input type="checkbox"/> 2 - Nej, ikke ændret	<input type="checkbox"/> 3 - Nedskåret			
11. Var produktionen begrænset eller hindret ultimo tællingskvartalet som følge af	61	<input type="checkbox"/> 1 - Arbejdskraftmangel	<input type="checkbox"/> 2 - Utilstrækkelig kapacitet				
	61	<input type="checkbox"/> 3 - Utilstrækkelig efterspørgsel	<input type="checkbox"/> 4 - Ingen produktionsbegrænsninger				

Figur 1. Uddrag af spørgeskema til indsamling af oplysningerne vedrørende »Konjunkturbarometret for industrien«.

Note: Ved vurderingen af det faktiske forløb sammenholdes indeværende kvartal med forrige kvartal. Tilsvarende vurderes det forventede forløb i forhold til indeværende kvartal.

tidspunkter og afdækker derved, hvorledes nettoforskydningerne er fremkommet som resultatet af en række individuelle skift.¹ For det andet er karakteren af de stillede spørgsmål bemærkelsesværdig. Figur 1 gengiver et uddrag af det anvendte spørgeskema. Det fremgår heraf, at virksomhederne dels bedes vurdere den forløbne

1. Et velkendt eksempel til illustration af forskellen mellem brutto- og nettoforskydninger vil være de politiske »indeks«, som jævnligt udarbejdes af flere analyseinstitutter. Disse består typisk af en opgørelse af vælgertilslutningen til de enkelte partier, målt marginalt. Ved sammenligning med forrige »indeks« eller med stemmefordelingen ved sidste folketingsvalg kan eventuelle ændringer noteres. Der er her tale om nettoforskydninger, idet man ikke kan afgøre, hvorledes disse er sammensat, dvs. hvor en tilgang til et parti kommer fra, og hvor en afgang går hen. For at måle dette må de samme vælgere indgå i de to undersøgelser, som sammenholdes, altså udgøre et panel. (Man prøver undertiden at omgå dette krav ved at spørge om stemmeafgivningen ved forrige valg, men denne tekniks anvendelighed er klart begrænset).

Tabel 1. Konjunkturbarometer for industri: Faktisk forløb i januar kvartal 1981 i forhold til oktober kvartal 1980 sammenholdt med det forventede forløb (tallene i parentes).

	Investeringsgodeindustri		Mellemproduktindustri		Forbrugsgodeindustri		Industri i alt				
	Mindre	Omtrent uændret	Mindre	Omtrent uændret	Mindre	Omtrent uændret	Mindre	Omtrent uændret			
Produktion	44 (32)	42 (51)	14 (17)	42 (34)	23 (16)	25 (38)	54 (50)	21 (12)	36 (35)	45 (50)	19 (15)
Beskæftigelse ultimo kvartalet	44 (41)	44 (50)	12 (9)	41 (33)	16 (5)	33 (24)	55 (71)	12 (5)	39 (32)	48 (62)	13 (6)
Samlet ordreindgang	39 (31)	39 (50)	22 (19)	39 (28)	25 (15)	34 (32)	40 (51)	26 (17)	37 (31)	38 (52)	25 (17)
Samlede løbende investeringer	26 (18)	64 (66)	10 (16)	34 (31)	8 (9)	21 (15)	72 (77)	7 (8)	26 (20)	66 (69)	8 (11)
Investeringsplaner ændret i kvartalet	13 (.)	84 (.)	3 (.)	12 (.)	14 (.)	5 (.)	91 (.)	4 (.)	9 (.)	84 (.)	7 (.)

Kilde: Statistiske Efterretninger, B 23, 1981, p. 538.

udvikling på en række områder i forhold til forrige kvartal samt forventningerne til næste kvartal, og dels angive, hvorledes en række størrelser af statusmæssig art forholder sig til det »normale« eller »tilstrækkelige« niveau. Det er karakteristisk for de stillede spørgsmål, at de kræver svar, der er udtrykt i *kvalitative* termer, f.eks. »større«, »uændret« og »mindre« eller »mere end tilstrækkelig«, »tilstrækkelig« og »mindre end tilstrækkelig«.

Denne forenkede spørgeform, som bygger på en metode, udviklet af det tyske IFO-institut, er bevidst valgt af hensyn til mulighederne for en hurtig svarafgivning. På den anden side må det erkendes, at metoden indebærer en del *måleusikkerhed* samt, at målingernes *sensitivitet* er reduceret i forhold til oplysninger af *kvantitativ* art.

»Konjunkturbarometeret« offentliggøres på en form, der afspejler datamaterialets kvalitative karakter. Tabel 1 viser et typisk eksempel vedrørende det faktiske og det forventede udviklingsforløb for produktion, beskæftigelse, ordreindgang, løbende investeringer og ændringer i investeringsplaner. Den viser svarfordelingen på de tre svarkategorier »mindre«, »uændret« og »større« for de 5 spørgsmål. Der er forinden foretaget en vægtning af virksomhederne i forhold til deres andel af den samlede beskæftigelse i industrien. Desuden er der sket en opdeling efter arten af den frembragte produktion. Tabellen skal derfor læses således, at der eksempelvis var virksomheder, repræsenterende 44 % af de beskæftigede inden for investeringsgodeindustrien, der vurderede produktionen som »mindre« i det pågældende kvartal i forhold til foregående kvartal.² Tilsvarende tabeller publiceres, opdelt på 29 industrigrupper og på amtskommuner.

Som »Konjunkturbarometeret« fremtræder, er det klart, at dets anvendelse må knytte sig til svarfordelingen på de enkelte spørgsmål, *taget hver for sig*. Der er derimod ingen mulighed for at vurdere fordelingen på to eller flere spørgsmål samtidigt med henblik på en analyse af deres eventuelle sammenhæng. Dette er beklageligt, fordi almindelig økonomisk indsigt tilsiger, at sådanne sammenhænge vil kunne forventes. De fire spørgsmål i fig. 1 om udviklingsforløbet for et kvartal vil således alle afspejle nogle generelle træk ved den (faktiske eller forventede) økonomiske udvikling. Det er netop denne egenskab, som er baggrund for deres anvendelse i barometeret som *konjunkturindikatorer*.

Dernæst bemærkes det, at datamaterialets karakter af paneldata ikke udnyttes med den valgte publiceringsform. Et interessant aspekt ved materialet vil være graden af overensstemmelse mellem virksomhedernes konjunkturforventninger og den faktiske

2. Som anført på skemaet i fig. 1 bliver virksomhederne bedt om at se bort fra eventuelle sæsonmæssige svingninger.

udvikling, men en vurdering heraf alene på grundlag af nettoforskydninger (de marginale fordelinger) som i tabel 1 vil som tidligere nævnt kun give et ufuldstændigt billede af de bruttobevægelser, som kan have fundet sted. En analyse heraf kræver kendskab til de simultane fordelinger af spørgsmålene vedrørende den faktiske og forventede udvikling, og dette forudsætter udnyttelse af materialets panelegenskaber.

Det synes at være en almindelig opfattelse blandt økonomer og andre brugere af »Konjunkturbarometeret«, at de indsamlede oplysninger med deres særegne karakter først og fremmest er anvendelige i konjunkturbedømmelsen i kraft af deres aktualitet, men i mindre grad egner sig til egentlige systematiske analyser.³ Den mest omfattende undersøgelse af danske konjunkturdata af den foreliggende type er Munksgårds afhandling (Munksgård, 1974), som bygger på et materiale fra perioden 1956-59 vedrørende tekstilbranchen. Inden for rammerne af en driftsøkonomisk model formuleres og afprøves en række hypoteser om dataenes egenskaber og anvendelighed som led i en »konjunkturtest«. Afhandlingen bærer i nogen grad præg af at være pionerarbejde, og det må erkendes, at datagrundlaget i nogle sammenhænge er ganske spinkelt, og at de anvendte statistiske metoder (derfor) er elementære.

Denne artikel, som er et uddrag af et større arbejde,⁴ har til formål at demonstrere anvendelse af en egentlig statistisk model, den såkaldte *latente klasse (LK) model*, ved analyse af data fra »Konjunkturbarometeret«.⁵ Når denne model er fundet velegnet, skyldes det hovedsageligt følgende forhold:

- (1) den er formuleret med henblik på *kvalitative* data, hvorved vi vil forstå svar af den kategoriske type, enten uden ordning (ja/nej, mand/kvinde etc.) eller med ordning (større/uændret/mindre, i høj grad/i ringe grad etc.);
- (2) den tager udgangspunkt i *sammenhængen* mellem de enkelte spørgsmål, som opfattes som *indikatorer* for en bagvedliggende, generel konjunkturvariabel, hvis karakter det er et af modellens formål at afdække;
- (3) sammenhængen mellem indikatorer og bagvedliggende variable opfattes som *probabilistisk*;
- (4) den tillader en statistisk tilfredsstillende estimation og afprøvning, herunder tests af hypoteser vedrørende modellens parametre;
- (5) den er relativ simpel at fortolke.

3. Jf. Skak-Nielsens formulering i J. Hansen (red.): »Industrivirksomhedens Konjunkturvarslingsystem«, 1. udg. Industrirådet 1976.

4. Nogle mere omfattende analyser foreligger i form af en række arbejdsrapporter. De kan rekvireres ved henvendelse til forfatteren.

5. Efter færdiggørelse af denne artikel har forfatteren fået kendskab til arbejdet Flaig (1983), som tager det tilsvarende udgangspunkt i tyske data.

Vi skal i næste afsnit give en kort fremstilling af det begrebsmæssige grundlag for latent strukturanalyse. Dernæst vil de statistiske inferensproblemer i LK-modellen blive behandlet, hvorefter modellen illustreres på data vedrørende et enkelt kvartal. Vi har valgt at belyse den faktiske udvikling i 1. kvartal 1981 på grundlag af tre af de i fig. 1 indeholdte konjunkturindikatorer.

Det begrebsmæssige grundlag for latent strukturanalyse

LK-modellen er et specialtilfælde af en mere generel samling af modeller, som Lazarsfeld (1950, 1954) kaldte *latent strukturanalyse*, se tillige Lazarsfeld og Henry (1968) for en samlet fremstilling. Modellerne er beslægtede med den mere velkendte faktoranalyse, men adskiller sig dog på flere væsentlige punkter. Ligesom i faktoranalysen tager den latente strukturanalyse sit udgangspunkt i sondringen mellem de direkte observerbare eller *manifeste* observationer og de bagvedliggende eller *latente* faktorer, som de afspejler. De observerede variable ses som ufuldstændige indikatorer for disse faktorer, som i *statistisk* forstand »forklarer« sammenhængen mellem de observerede variable, dvs. hvis vi kunne kontrollere værdien af faktorerne, ville de manifeste variable være *statistisk uafhængige*.

Formålet med modellerne er at afdække de latente begrebs egenskaber og forbindelse til de manifeste variable. I faktoranalysen sker dette typisk ved at antage en lineær struktur for relationen mellem de manifeste og latente variable. Udgangspunktet er sædvanligvis en kovarians- eller korrelationsmatrix, hvilket implicit forudsætter, at alle målinger sker på (mindst) intervallskalaniveau. I *eksplorative* analyser vil faktorløsningen ikke være éntydig, hvilket leder til problemerne omkring *rotation* af faktorerne. Det grundlæggende princip er her at søge frem til en »simpel struktur«, dvs. en endelig løsning som *fortolkningsmæssigt* forekommer rimelig. Anvendes faktoranalysen *hypoteseafprovende* indføres a priori restriktioner på modellens parametre, hvorved faktorløsningen bliver éntydigt fastlagt. Dette er synspunktet hos Jöreskog (1969), se tillige tilføjes i Jöreskog og Sörbom (1979). Når en faktormodel er accepteret som tilfredsstillende, estimeres *faktorscores* som udtryk for de bedste skøn over de bagvedliggende sande faktorværdier og disse kan da anvendes i efterfølgende analyser.

Som Lazarsfeld bemærkede er det sjældent at støde på intervallskalerede målinger i samfundsvidenskaberne. Latent strukturanalyse bygger derfor på *kvalitative* indikatorer, dvs. variable, der er nominalt eller ordinalt skalerede. Som en konsekvens heraf er det grundlæggende mål for sammenhængen mellem de manifeste variable den *simultane sandsynlighedsfordeling* og der gøres ingen forudsætning om linearitet mellem variablerne på de to niveauer. Antages den latente variabel også at være *kvalitativ*, bestående af kategorier med eller uden orden, fremtræder den latente

klasse (LK) model. Den latente klassevariabel »forklarer« her sammenhængen mellem indikatorerne i den forstand, at der *inden for* hver klasse er *statistisk uafhængighed*. Når modellen estimeres effektivt, f.eks. ved *maximum likelihood* princippet som nedenfor, opstår der ikke noget rotationsproblem, og den latente struktur kan vurderes direkte ved betragtning af (1) den relative fordeling af den latente variabel på klasser, dvs. klassestørrelsen, og (2) de betingede fordelinger for hver klasse på de manifesterede indikatorers kategorier. Endelig kan der svarende til begrebet faktorscores beregnes *klassetilhørsandsynligheder*, der viser den betingede sandsynlighed for at tilhøre en vilkårlig klasse, når niveauerne på indikatorerne er givne.

Trods betydningen af de centrale ideer bag latent strukturanalyse, er det først med fremkomsten af moderne datamaskiner, at modellerne i praksis er blevet tilgængelige for analytikere. På grundlag af en række arbejder af Goodman (1974a, 1974b, 1975a), som anviste en algoritme til bestemmelse af maximum likelihood estimater af LK-modellens parametre, har Clogg (1977) udviklet et meget fleksibelt edb-program, MLLSA (for *maximum likelihood latent structure analysis*), der kan behandle en lang række LK-modeller med eller uden restriktioner på parametrene. Alle beregninger i denne artikel er udført ved hjælp af MLLSA-programmet.

Statistisk inferens i latent klasseanalyse

Efter at have opridset det begrebsmæssige grundlag for latent struktur analyse skal vi her give en mere formaliseret definition af det specialtilfælde, der kaldes LK-modellen og kort påpege nogle af modellens statistiske egenskaber.

Definition. Uden tab af generalitet vil vi betragte tre kvalitative manifesterede variable A , B og C med niveauerne $i = 1, \dots, I$, $j = 1, \dots, J$ og $k = 1, \dots, K$. $\{ABC\}$ angiver den tredimensionale tabel, der fremkommer ved at krydse de tre variable med hinanden. Tabellen vil indeholde $I \cdot J \cdot K$ celler og den typiske celle betegnes (i, j, k) . Det totale antal observationer (respondenter) til fordeling i tabellen er N og antallet af observationer i den typiske celle kaldes x_{ijk} . Sandsynligheden for en observation i (i, j, k) er π_{ijk} . Vi antager, at fordelingen af de N observationer på tabellens celler kan beskrives ved en multinomisk fordeling med sandsynlighedsparametre π_{ijk} .

Lad nu X være en kvalitativ latent variabel med niveauer $t = 1, \dots, T$. Hver observation tilhører en og kun én klasse t af X , dvs.

$$\pi_{ijk} = \sum_t \pi_{ijkt}^{ABCX} \quad (1)$$

hvor π_{ijkt}^{ABCX} angiver sandsynligheden for en observation i den typiske celle (i, j, k, t) i

{*ABCX*}-tabellen. Bemærk, at {*ABCX*} ikke er direkte observerbar, da *X* er latent. Antag dernæst, at *X* i statistisk forstand »forklarer« sammenhængen mellem de manifeste variable, dvs. når *X* fastholdes eller kontrolleres bliver variablerne *A*, *B* og *C* (betinget) uafhængige. Formelt kan dette udtrykkes:

$$\pi_{ijkl}^{ABCX} = \pi_i^X \cdot \pi_{it}^{\bar{A}X} \cdot \pi_{jt}^{\bar{B}X} \cdot \pi_{kt}^{\bar{C}X} \quad (2)$$

hvor $\pi_{it}^{\bar{A}X}$ er den betingede sandsynlighed for at observere *A* på niveau *i*, når *X* er givet på niveau *t*, og $\pi_{jt}^{\bar{B}X}$ og $\pi_{kt}^{\bar{C}X}$ er defineret analogt. π_i^X er den marginale fordeling af *X*, som vi vil kalde klassestørrelsen. Da alle parametre i (1) og (2) er sandsynligheder må følgende lineære bånd gælde:

$$\sum_i \pi_i^X = 1 \quad (3a)$$

$$\sum_i \pi_{it}^{\bar{A}X} = \sum_j \pi_{jt}^{\bar{B}X} = \sum_k \pi_{kt}^{\bar{C}X} = 1 \quad (3b)$$

Kombineres (1) og (2) er LK-modellen defineret ved parametriseringen

$$\pi_{ijk} = \sum_{t=1}^T \pi_i^X \cdot \pi_{it}^{\bar{A}X} \cdot \pi_{jt}^{\bar{B}X} \cdot \pi_{kt}^{\bar{C}X} \quad (4)$$

sammen med restriktionerne (3a) og (3b). LK-modellen tilhører således klassen af *parametriske, multinomiske fordelinger* for hvilken der findes generelle teoretisk-statistiske resultater, se Rao (1973) eller Andersen (1980a).

Estimation. Med udgangspunkt i en given antaltabel (x_{ijk}) vil estimationen af en LK-model omfatte

(I) antallet af klasser *T*.

(II) for et givet antal klasser, estimation af det *fundamentale parametersæt*:

(i) klassestørrelserne π_1^X, \dots, π_T^X , dvs. ialt $(T-1)$ frie parametre under hensyntagen til (3a)

(ii) de betingede sandsynligheder

$$\pi_{1t}^{\bar{A}X}, \dots, \pi_{It}^{\bar{A}X}, \quad t = 1, \dots, T$$

$$\pi_{1t}^{\bar{B}X}, \dots, \pi_{Jt}^{\bar{B}X}, \quad t = 1, \dots, T$$

$$\pi_{1t}^{\bar{C}X}, \dots, \pi_{Kt}^{\bar{C}X}, \quad t = 1, \dots, T$$

dvs. ialt $T \cdot (I + J + K - 3)$ frie parametre, når (3b) iagttages.

Antallet af klasser T i (I) er ofte ikke kendte a priori og må da bestemmes eksplorativt ved udnyttelse af resultater i (II). Estimationen i (II) sker ved maximum likelihood (ML) princippet, som kan vises at være *asymptotisk efficient*, dvs. når antallet af observationer N bliver (meget) stort kan vi være sikre på at få præcise og pålidelige estimater.

Goodman (1974a) har påvist, hvorledes ML estimatorerne kan beregnes ved hjælp af en iterativ procedure, som nu kendes under betegnelsen EM-algoritmen, se Andersen (1980b) og Poulsen (1981). Parameterløsningen kan dog ikke vises at være entydig, og det er derfor tilrådeligt at gentage iterationerne fra et andet udgangspunkt og i tilfælde af flere løsninger vælge den, der svarer til den største værdi af likelihoodfunktionen. Tilfælde af løsninger på randen 0 eller 1 af parameterområdet behandles lettest ved at opfatte disse som a priori restriktioner på LK-parametrene, som beskrevet nedenfor i afsnittet om betinget testning.

Vurdering af »goodness-of-fit«. Når skøn over de fundamentale parametre i LK-modellen er opnået, kan modellens kvalitet som beskrivende tabellen $\{ABC\}$ vurderes ved den velkendte Pearson X^2 -størrelse:

$$\begin{aligned} X^2 &= \sum_{\text{alle celler}} \frac{(\text{observeret-forventet})^2}{\text{forventet antal}} \\ &= \sum_{i,j,k} \frac{(x_{ijk} - N \cdot \hat{\pi}_{ijk})^2}{N \cdot \hat{\pi}_{ijk}} \end{aligned} \quad (5)$$

hvor $\hat{\pi}_{ijk}$ er skøn over π_{ijk} på grundlag af (4) med ML-estimerne for LK-parametrene indsat. En anden »goodness-of-fit« indikator er likelihood-ratio størrelsen

$$\begin{aligned} G^2 &= 2 \sum_{\text{alle celler}} (\text{observeret}) \cdot \log_e \frac{\text{observeret}}{\text{forventet}} \\ G^2 &= 2 \sum_{i,j,k} x_{ijk} \cdot \log_e \frac{x_{ijk}}{N \cdot \hat{\pi}_{ijk}} \end{aligned} \quad (6)$$

Hvis LK-modellen er *sand og identificeret*⁶ vil X^2 og G^2 have samme asymptotiske χ^2 -fordeling med et antal frihedsgrader df , der bestemmes ved

6. En LK-model siges at være identificeret, såfremt der til et sæt af multinomiske sandsynligheder (π_{ijk}) findes præcis ét sæt af LK-parametre. Se iøvrigt diskussionen hos Goodman (1974b).

$$df = \text{antal celler i } \{ABC\}\text{-tabellen} - 1 \\ - \text{antal fundamentale parametre}$$

Dvs.

$$df = I \cdot J \cdot K - 1 - (T - 1) - T(I + J + K - 3) \quad (7)$$

i den generelle T -klasse model med tre manifeste variable. Sædvanligvis er forskellen mellem X^2 - og G^2 -værdierne underordnet, såfremt modellen holder, men G^2 er snævert forbundet med ML-estimationsprincippet, idet G^2 herved minimeres. Desuden har størrelsen vigtige egenskaber i forbindelse med sammenligning af hierarkiske modeller.

Betingede tests. To LK-modeller M_1 og M_2 med samme antal klasser T , siges at være hierarkisk forbundne, såfremt den ene model, f.eks. M_1 , indeholder den anden, M_2 , som et specialtilfælde, idet en eller flere fundamentale parametre i M_1 er båndlagt i M_2 . To typer af bånd er værdirestriktioner, hvor en parameters værdi er fast, f.eks. $\pi_1^X = 0.5$ eller $\pi_{11}^{\bar{A}X} = 0.25$, og lighedsrestriktioner, hvor nogle parametre er fastsat til samme (ukendte) værdi, f.eks. $\pi_1^X = \pi_2^X$ og $\pi_{11}^{\bar{A}X} = \pi_{12}^{\bar{B}X} = \pi_{13}^{\bar{C}X}$. Det kan vises, at et test for de ved M_2 indførte restriktioner, betinget af M_1 , kan fås ved at betragte differencen:

$$G^2(M_2; M_1) = G^2(M_2) - G^2(M_1) \quad (8)$$

som hvis M_2 er sand vil have en asymptotisk χ^2 -fordeling med antal frihedsgrader

$$df(M_2; M_1) = df(M_2) - df(M_1) \\ = \text{antal restriktioner indført ved } M_2 \quad (9)$$

Tilordning til klasser. Når en LK-model er blevet accepteret som velegnet til beskrivelse af data, har det ofte stor interesse at undersøge sammenhængen mellem den latente klassevariabel X og andre målte kendetegn ved analyseenhederne (respondenterne). Til dette formål anvendes Bayes' sætning til beregning af de posteriore tilhørsandsynligheder,⁷ givet observationerne (ijk) på de manifeste variable:

$$\pi_{ijkt}^{ABC\bar{X}} = \pi_{ijkt}^{ABCX} / \pi_{ijk} \\ = \pi_t^X \cdot \pi_{it}^{\bar{A}X} \cdot \pi_{jt}^{\bar{B}X} \cdot \pi_{kt}^{\bar{C}X} / \pi_{ijk} \quad (10)$$

Ved indsættelse af estimater for LK-parametrene og anvendelse af (4) kan skøn

7. Lazarsfeld kaldte $\pi_{ijkt}^{ABC\bar{X}}$ for »recruitment probabilities«.

over $\pi_{ijkl}^{ABC\bar{X}}$ beregnes. Dernæst kan analyseenhederne tilordnes klassen med den største tilhørsandsynlighed, idet det kan vises, at denne procedure minimerer det forventede antal fejlklassificeringer, E , beregnet ved formlen:

$$E = \sum_{i,j,k} x_{ijk} \cdot (1 - \pi_{ijkl}^{ABC\bar{X}}) \quad (11)$$

hvor

$$\pi_{ijkl}^{ABC\bar{X}} = \max_t \{ \pi_{ijkl}^{ABC\bar{X}} \}$$

Anvendelse af LK-modellen på »Konjunkturbarometeret for industrien«

Af de formelle udledninger ovenfor fremgår det, at estimation af LK-modellen ved ML-metoden kræver kendskab til den flerdimensionale antalstabel $\{ABC\}$, dvs. oplysninger på et væsentligt mere detaljeret niveau end den form, hvorpå resultaterne løbende publiceres. Det har derfor været nødvendigt at lave et udtræk af grundmaterialet, der er opbevaret i Danmarks Statistik. Af hensyn til virksomhedernes anonymitet blev en række identifikationsvariable slettet, inden udlevering af materialet. Det skal desuden bemærkes, at virksomhederne i denne analyse indgår *uvægtet* uden hensyntagen til beskæftigelsesomfang. Der er dog i princippet intet til hinder for at vejede data kunne have været anvendt.

LK-modeller uden a priori restriktioner. Som grundlag for analysen har vi valgt data vedrørende udviklingen i 1. kvartal 1981 sammenlignet med forrige kvartal. Vi vil betragte følgende tre konjunkturindikatorer:

- A: Faktisk produktion
- B: Faktisk beskæftigelse
- C: Faktisk udvikling i ordreindgangen

Alle tre variable er trikotome med niveauerne »1« = »mindre«, »2« = »uændret« og »3« = »større«. Databasen omfatter $N = 497$ respondenter og den grundlæggende antalstabel $\{ABC\}$ er gengivet som kolonne (3) i tabel 2.

Bemærk, at 6 celler i tabellen indeholder 0 respondenter. Det betyder, at den asymptotiske teoris resultater kun med tilnærmelse kan antages at gælde. Desuden kan der opstå problemer med celler, som under nogle LK-modeller har *forventede* antal lig nul, hvilket ville kræve, at antallet af frihedsgrader i (7) korrigeres tilsvarende. Vi har valgt her *ad hoc* at lægge værdien 0.5 til alle celler med nul.⁸

8. Denne fremgangsmåde kendes fra kontingenstabellanalyse ved log-lineære modeller, se f.eks. Goodman's ECTA program, Goodman (1975b).

Tabel 2. Den grundlæggende antalstabel $\{ABC\}$ og nogle resultater for den tilpassede 3-klassemodel.

Celle nr.	Svar-mønster			Observeret antal	Forventet antal	Modal klasse ssh.	Tilordnet klasse
	i	j	k	x_{ijk}	$N \cdot \hat{\pi}_{ijk}$	$\hat{\pi}_{ijk}^{ABC}$	t^*
1	1	1	1	84	84.26	0.975	1
2	2	1	1	21	21.23	0.798	2
3	3	1	1	2	1.84	0.578	1
4	1	2	1	33	32.44	0.813	1
5	2	2	1	50	49.99	0.972	2
6	3	2	1	2	4.08	0.916	3
7	1	3	1	0	0.59	0.867	1
8	2	3	1	0	6.40	0.869	2
9	3	3	1	5	3.16	0.998	3
10	1	1	2	16	16.43	0.787	1
11	2	1	2	27	28.70	0.976	2
12	3	1	2	4	2.67	0.937	3
13	1	2	2	15	14.24	0.699	2
14	2	2	2	82	80.69	0.996	2
15	3	2	2	13	12.14	0.995	3
16	1	3	2	0	0.27	0.428	3
17	2	3	2	0	0.67	0.868	2
18	3	3	2	8	10.19	1.000	3
19	1	1	3	8	7.49	0.860	1
20	2	1	3	10	8.02	0.949	2
21	3	1	3	8	9.35	0.991	3
22	1	2	3	4	5.28	0.512	2
23	2	2	3	21	22.32	0.979	2
24	3	2	3	46	44.81	0.999	3
25	1	3	3	0	0.48	0.876	3
26	2	3	3	0	0.47	0.661	3
27	3	3	3	38	37.76	1.000	3

Anm. Til alle celler med 0 er 0.5 adderet inden beregningen af det forventede antal observationer.

Derved bliver $N = 500$. De strukturelle parametre er ganske robuste over for denne korrektion.

Vi afprøver nu eksplorativt modeller med $T = 1, 2$ og 3 klasser. De tilhørende G^2 -værdier og deres signifikans fremgår af tabel 3.

For $T = 1$ er LK-modellen M_1 identisk med hypotesen om (global) uafhængighed mellem de tre indikatorer. Med $G^2(M_1) = 465.99$ og $df(M_1) = 20$ må vi klart forkaste denne hypotese. Heller ikke modellen med $T = 2$ klasser giver en tilfredsstillende beskrivelse af antalstabellen, men for $T = 3$ fås $G^2 = 4.95$, som med $df = 7$

Tabel 3. »Goodness-of-fit« for LK-modeller uden a priori restriktioner med $T=1, 2$ og 3 klasser.

Model M_i	Antal klasser T	$G^2(M_i)$	$df(M_i)$	p -niveau
M_1	1	465.99	20	0.000
M_2	2	127.79	13	0.000
M_3	3	4.951	7*	0.666

* Korrigeret for ét estimat på randen, jf. teksten.

Tabel 4. Estimeret 3-klasse model uden a priori restriktioner.
Faktisk udvikling, 1. kvartal 1981.

		Latent klasse		
		1	2	3
Klassestørrelse		0.287	0.460	0.253
Betingede sandsynligheder				
A: Produktion	mindre	0.939	0.110	0.011
	uændret	0.049	0.890	0.008
	større	0.012	0.000*	0.981
B: Beskæftigelse	mindre	0.753	0.257	0.101
	uændret	0.242	0.738	0.488
	større	0.005	0.005	0.411
C: Ordreindgang	mindre	0.809	0.322	0.062
	uændret	0.127	0.553	0.199
	større	0.064	0.145	0.739

* Estimeret randværdi

repræsenterer et glimrende fit. Antallet af frihedsgrader er opgjort til 7 og ikke de 6, som fås af formlen (7) for LK-modeller uden begrænsninger på parameterene. Det skyldes, at én estimeret parameter (π_{32}^{AX}) ligger på randen (0.0). Vi må derfor betragte denne restriktion som indført a priori, hvorved antallet af frihedsgrader forøges med 1. Den estimerede 3-klassemodel er gengivet i tabel 4.

Fortolkningen af den afdækkede LK-struktur er forholdsvis indlysende. De tre klasser består af respondenter (virksomheder), som overvejende har samme udvikling (mindre, uændret eller større) på alle tre indikatorer. Klassen af »uændret«

omfatter næsten halvdelen af respondenterne, og de to øvrige er stort set lige store. Af de tre indikatorer er produktionsudviklingen den variabel, der klart adskiller de tre klasser. Tilsvarende er beskæftigelsesudviklingen den ringeste diskriminator. Dette er i sig selv interessante konklusioner, idet det viser, at mens de tre konjunkturindikatorer nok er koordinerede i udviklingen, er det primært produktionen som på kort sigt karakteriserer forskellene i konjunkturvurdering. En forklaring herpå kan være, at beskæftigelsen inden for den betragtede tidshorisont (kvartal) er mindre fleksibel end produktionen. Endelig er ordreindgangen i høj grad bestemt af eksogene faktorer og kan på kort sigt udvise betydelig variation.

På grundlag af den estimerede 3-klassemodel kan vi nu anvende den skitserede klassifikationsprocedure efter modalværdien af klasses tilhørs sandsynlighederne. Klassifikationen fremgår af tabel 2, kolonne 5 og 6. Til vurdering af klassificeringens kvalitet kan vi beregne det forventede antal fejlklassificeringer E iflg. (11). Vi får:

$$E = 470.99$$

dvs. kun 5,8 pct. af respondenterne kan forventes henført til den forkerte klasse.

3-klassemodellen med restriktioner. Som illustration af princippet om betinget testning skal vi nu indføre restriktioner i den generelle 3-klassemodel M_1 ovenfor. På baggrund af den estimerede struktur vil vi afprøve en hypotese om, at *produktionsvariablen er perfekt indikator for den latente variabel*. Formelt indebærer denne antagelse en éntydig korrespondance mellem variablene A og X , dvs.:

$$\pi_{1t}^{\bar{A}X} = \begin{cases} 1 & \text{for } t=1 \\ 0 & \text{for } t=2,3 \end{cases} \quad (12a)$$

$$\pi_{2t}^{\bar{A}X} = \begin{cases} 1 & \text{for } t=2 \\ 0 & \text{for } t=1,3 \end{cases} \quad (12b)$$

$$\pi_{3t}^{\bar{A}X} = \begin{cases} 1 & \text{for } t=3 \\ 0 & \text{for } t=1,2 \end{cases} \quad (12c)$$

Herved påføres ialt 6 uafhængige bånd, men da $\pi_{32}^{\bar{A}X} = 0$ i forvejen var gældende, har den restriktive model M_2 5 frihedsgrader flere end M_1 , dvs. $df(M_2) = 12$. Den estimerede model har $G^2(M_2) = 11.299$, hvilket i sig selv repræsenterer et godt fit. Det betingede test for gyldigheden af restriktionerne (12a)-(12c) giver:

$$G^2(M_2; M_1) = 11.299 - 4.951 = 6.348$$

$$df(M_2; M_1) = 12 - 7 = 5$$

hvilket ikke kan anses for signifikant.

Tabel 5. Estimeret 3-klasse model med produktionsudviklingen (*A*) som perfekt indikator.

		Latent klasse		
		1	2	3
Klassestørrelse		0.323	0.425	0.252
Betingede sandsynligheder:				
A: Produktion	mindre	1.00†	0.00†	0.00†
	uændret	0.00†	1.00†	0.00†
	større	0.00†	0.00†	1.00†
B: Beskæftigelse	mindre	0.669	0.273	0.111
	uændret	0.322	0.720	0.484
	større	0.009	0.007	0.405
C: Ordreindgang	mindre	0.728	0.337	0.071
	uændret	0.195	0.515	0.198
	større	0.077	0.148	0.730

† Fikseret a priori.

Den estimerede, restriktive model fremgår af tabel 5.

Da den latente klassevariabel *X* her er sammenfaldende med den ene manifesterede variabel *A*, produktionsudviklingen, vil klassestørrelsen naturligvis være identisk med den marginale fordeling af *A*. De øvrige parameterestimer har kun ændret sig lidt i forhold til modellen uden a priori restriktioner.

For at illustrere anvendelsen af lighedsrestriktioner skal vi nu afprøve en hypotese om *symmetri* i svarsandsynlighederne for klasse 1 og klasse 3. Vi indfører de fire bånd:

$$\pi_{11}^{BX} = \pi_{33}^{BX} \quad (13a)$$

$$\pi_{31}^{BX} = \pi_{13}^{BX} \quad (13b)$$

$$\pi_{11}^{CX} = \pi_{33}^{CX} \quad (13c)$$

$$\pi_{31}^{CX} = \pi_{13}^{CX} \quad (13d)$$

mens lighederne $\pi_{21}^{BX} = \pi_{23}^{BX}$ og $\pi_{21}^{CX} = \pi_{23}^{CX}$ automatisk følger af de lineære bånd (3b). Denne endnu mere restriktive model M_3 vil derfor have 4 frihedsgrader flere end M_2 . Estimation af modellen giver $G^2(M_3) = 40.188$ som med $df = 16$ repræsenterer et dårligt fit. Det af M_2 betingede test giver ligeledes anledning til afvisning af symmetrihypotesen, idet

$$G^2(M_3; M_2) = 40.188 - 11.299 = 28.889$$

med $df = 4$ frihedsgrader er klart signifikant.

Med accepten af produktionsvariablen som perfekt indikator har vi en vis forstand overflødiggjort den latente variabel ved beskrivelsen af datamaterialet. Det forventede antal fejlklassificeringer E er således reduceret til nul, da tilordningen til klasser nu ikke længere er behæftet med usikkerhed. Vi kan konkludere, at en model med betinget uafhængighed, hvor produktionsvariablen A »forklarer« sammenhængen mellem beskæftigelsen B og ordreindgangen C , giver en god sammenfatning af $\{ABC\}$ -tabellen. Cellesandsynlighederne π_{ijk} kan med den indførte notation i dette tilfælde skrives:

$$\pi_{ijk} = \pi_i^A \cdot \pi_{ij}^{AB} \cdot \pi_{ik}^{AC} \quad (14)$$

Andre metoder, f.eks. direkte beregning af skøn over parametrene i (14) eller specifikation af den tilhørende log-lineære model,⁹ kunne have været anvendt. En række log-lineære modeller har således en ækvivalent repræsentation som restriktive LK-modeller. Det drejer sig om strukturer som indeholder en form for uafhængighed (betinget eller ubetinget) og som derfor er forholdsvis lette at fortolke. Mere komplicerede log-lineære modeller vil ikke kunne skrives på LK-form, men er samtidig vanskelige at begribe intuitivt. I LK-modellen er alle parametre derimod umiddelbart forståelige som sandsynligheder og desuden omfattes modeller, som ikke er log-lineære. Det ville iøvrigt være vanskeligt at afprøve restriktioner af typen (13a)-(13d) med den log-lineære formulering, mens det er problemfrit i LK-modellen.

Endelig er den måske vigtigste forskel erkendelsen af, at observationer i samfundsvidenskaberne som regel må opfattes som ufuldstændige indikatorer for en bagvedliggende struktur, som kan være vanskelig at afdække, men som til gengæld vil være mere stabil og derfor af større interesse, teoretisk såvel som praktisk.

Afsluttende bemærkninger

I denne artikel har vi foreslået anvendelsen af en specifik, statistisk model til beskrivelse af et datamateriale, som produceres og publiceres løbende af Danmarks Statistik. Formålet har blandt andet været mere principielt at pege på mulighederne for at præsentere resultaterne på en form, som ville gøre dem mere overskuelige og dermed anvendelige for brugerne. Dette kan som her ske ved hjælp af en teoretisk-statistisk model, ved anvendelsen af grafiske metoder eller ved andre datareducerende teknikker. Behovet såvel som mulighederne herfor afspejler den teknologiske udvikling på databehandlingsområdet. Moderne datateknik har på den ene side øget mulighederne for billigt og tilgængeligt at opbevare grundoplysningerne fra

9. En generel indføring i den log-lineære analysemetodik findes i Andersen (1980a).

statistikproduktionen, men samtidig er behovet for en meningsfyldt sammenfatning af data blevet presserende. På dette felt må bureaustatistikeren og den matematisk-orienterede statistiker kunne mødes. Muligheden for – mod passende vederlag – at kunne få lavet særkørsler af grundmaterialer i Danmarks Statistik er et første skridt på vejen i retning af en mere brugerorienteret statistikproduktion. Publicering af standardanalyser og konsulentbistand til løsning af individuelle analysebehov kunne være det næste.

Litteratur

- Andersen, E. B. 1980a. *Discrete Statistical Models With Social Science Applications*. Amsterdam: North Holland Publishing Co.
- Andersen, E. B. 1980b. *Latent Structure Analysis*. Research report No. 64. Statistical Institute, University of Copenhagen.
- Clogg, C. C. 1977. Unrestricted and Restricted Maximum Likelihood Latent Structure Analysis: A Manual for Users. Working Paper No. 1977-09, Pennsylvania State University.
- Flaig, G. 1983. *Die Analysen von Qualitativen Variablen mit Latenten-Struktur-Modellen*. Inaugural-dissertation, Universität Mannheim.
- Goodman, L. A. 1974a. The Analysis of Systems of Qualitative Variables When Some of the Variables are Unobservable. Part I: A Modified Latent Structure Approach. *American Journal of Sociology*, Vol. 79, No. 5. March.
- Goodman, L. A. 1974b. Exploratory Latent Structure Analysis. *Biometrika*, Vol. 61, No. 2, pp. 215-231.
- Goodman, L. A. 1975a. A New Model for Scaling Response Patterns: An Application of the Quasi-Independence Concept. *Journal of American Statistical Association*, 70, pp. 755-68.
- Goodman, L. A. 1975b. ECTA Program. Description for Users. Upubliceret manuskript. Chicago.
- Jöreskog, K. G. 1969. A General Approach to Confirmatory Factor Analysis. *Psychometrika*, Vol. 34, No. 2. June.
- Jöreskog, K. G. and D. Sörbom. 1979. *Advances in Factor Analysis and Structural Equation Models*. Cambridge, Mass.: Abt. Books.
- Lazarsfeld, P. F. 1950. The Logical and Mathematical Foundation of Latent Structure. I: S. A. Stouffer et al. (eds.): *Measurement and Prediction*. Princeton: Princeton University Press.
- Lazarsfeld, P. F. 1954. A Conceptual Introduction to Latent Structure Analysis. I: P. F. Lazarsfeld (ed.): *Mathematical Thinking in the Social Sciences*. New York: The Free Press.
- Lazarsfeld, P. F. and N. W. Henry. 1968. *Latent Structure Analysis*. Boston: Houghton Mifflin Publishing Co.
- Munksgård, H. 1974. *Konjunkturtesten. Et led i virksomhedens søge-læreproces*. København: Nyt Nordisk Forlag.
- Poulsen, C. S. 1982. *Latent Structure Analysis With Choice Modeling Applications*. Ph.D. dissertation, Wharton School, University of Pennsylvania. Udgivet ved Handelshøjskolen i Århus, skriftserie H, nr. 3.
- Rao, C. R. 1973. *Linear Statistical Inference and Its Applications*. New York: John Wiley and Sons.

statistikproduktionen, men samtidig er behovet for en meningsfyldt sammenfatning af data blevet presserende. På dette felt må bureaustatistikeren og den matematisk-orienterede statistiker kunne mødes. Muligheden for – mod passende vederlag – at kunne få lavet særkørsler af grundmaterialer i Danmarks Statistik er et første skridt på vejen i retning af en mere brugerorienteret statistikproduktion. Publicering af standardanalyser og konsulentbistand til løsning af individuelle analysebehov kunne være det næste.

Litteratur

- Andersen, E. B. 1980a. *Discrete Statistical Models With Social Science Applications*. Amsterdam: North Holland Publishing Co.
- Andersen, E. B. 1980b. *Latent Structure Analysis*. Research report No. 64. Statistical Institute, University of Copenhagen.
- Clogg, C. C. 1977. Unrestricted and Restricted Maximum Likelihood Latent Structure Analysis: A Manual for Users. Working Paper No. 1977-09, Pennsylvania State University.
- Flaig, G. 1983. *Die Analysen von Qualitativen Variablen mit Latenten-Struktur-Modellen*. Inaugural-dissertation, Universität Mannheim.
- Goodman, L. A. 1974a. The Analysis of Systems of Qualitative Variables When Some of the Variables are Unobservable. Part I: A Modified Latent Structure Approach. *American Journal of Sociology*, Vol. 79, No. 5. March.
- Goodman, L. A. 1974b. Exploratory Latent Structure Analysis. *Biometrika*, Vol. 61, No. 2, pp. 215-231.
- Goodman, L. A. 1975a. A New Model for Scaling Response Patterns: An Application of the Quasi-Independence Concept. *Journal of American Statistical Association*, 70, pp. 755-68.
- Goodman, L. A. 1975b. ECTA Program. Description for Users. Upubliceret manuskript. Chicago.
- Jöreskog, K. G. 1969. A General Approach to Confirmatory Factor Analysis. *Psychometrika*, Vol. 34, No. 2. June.
- Jöreskog, K. G. and D. Sörbom. 1979. *Advances in Factor Analysis and Structural Equation Models*. Cambridge, Mass.: Abt. Books.
- Lazarsfeld, P. F. 1950. The Logical and Mathematical Foundation of Latent Structure. I: S. A. Stouffer et al. (eds.): *Measurement and Prediction*. Princeton: Princeton University Press.
- Lazarsfeld, P. F. 1954. A Conceptual Introduction to Latent Structure Analysis. I: P. F. Lazarsfeld (ed.): *Mathematical Thinking in the Social Sciences*. New York: The Free Press.
- Lazarsfeld, P. F. and N. W. Henry. 1968. *Latent Structure Analysis*. Boston: Houghton Mifflin Publishing Co.
- Munksgård, H. 1974. *Konjunkturtesten. Et led i virksomhedens søge-læreprøces*. København: Nyt Nordisk Forlag.
- Poulsen, C. S. 1982. *Latent Structure Analysis With Choice Modeling Applications*. Ph.D. dissertation, Wharton School, University of Pennsylvania. Udgivet ved Handelshøjskolen i Århus, skriftserie H, nr. 3.
- Rao, C. R. 1973. *Linear Statistical Inference and Its Applications*. New York: John Wiley and Sons.