Erhvervsøkonomiske kommentarer

Af Holger Reenberg*)

Mindste kvadraters metode til analysering af tidsserier

Indledning

Bent Vabø gav mindste kvadraters metode (MKM) en ilde medfart i en artikel i dette tidsskrift (nr. 3, 1977). Hovedkonklusionen i denne artikel var, at MKM er uanvendelig til prognosticering, når de analyserede tidsserier er under indflydelse af sæsonvariationer og cykliske bevægelser.

*) Videnskabelig assistent, cand. agro., Det landøkonomiske Driftsbureau.

Jeg skal i det følgende vise, at den principielle fejl, som Bent Vabø »påviser« ved MKM, ikke skyldes mangler ved metoden, men derimod en ukvalificeret anvendelse af den. Min hensigt er dels at imødegå Bent Vabø's behandling af metoden dels at antyde de muligheder, der er knyttet til MKM.
Som dataeksempel anvendes det samme, som er anvendt i ovennævnte artikel. Herved ses der bort fra cykliske bevægelser og stokatiske elementer. Det skal dog understreges, at selv om der indføres et stokastisk element, ændres intet ved proceduren i det følgende, forudsat at der ikke forekommer seriekorrelation o.l.

MKM til simultan bestemmelse af trend og sæsoneffekt

Eksemplet, som anvendes i beregningerne, er anført i tabel 1.

Kvartal (T)	Salg (S)	Sæson (D)	Sæson-trend (H)
1	20	1	1
2	10	0	0
3	10	0	0
4	10	0	0
5	22	1	5
6	11	0	0
7	11	0	0
8	11	0	0

Tabel 1. Dataeksempel.

Man bedes specielt bemærke kolonne 3, som indeholder værdien på en dummyvariabel. Denne variabel indføres for at udnytte den information, der foreligger om sæsonvariationens karakter. I det alminde-

lige tilfælde, hvor sæsonudsvingene ikke fremstår så klart som i eksemplet, vil man starte med tre dummyvariable og siden udelade dem, som ikke bidrager med signifikante koefficienter.
Hvis man ikke på forhånd kan udelukke systematiske udsving i en tidsserie, vil det være halsløs gerning at forklare denne udelukkende som resultat af en konstant og en trendkomponent. De fleste økonometriske lærebøger understreger, at man vil få et »biased« skøn for trenden, såfremt der indgår systematiske variationer i tidsserien, og disse ikke indbygges i modellen.
I det aktuelle tilfælde må den simple model ($S=a+bT$) udvides med en sæsonvariabel, således at vi får:

(1) $S=a + bT + c_1D$, ($D=1$ for første-kvartalerne og $D=o$ ellers)

Koefficienten c_1 angiver den gennemsnitlige sæsoneffekt, og b er trenden renset for sæson.
Hvis afvigelserne mellem de faktiske værdier og middelværdien betegnes med små bogstaver (eksempel: $s=S - \bar{S}$), kan koefficienterne a, b og c_1 beregnes af følgende ligningssystem:

$$\begin{aligned} \Sigma\ st &= b\ \Sigma\ t^2 + c_1\ \Sigma\, dt \\ \Sigma\ sd &= b\ \Sigma\ dt + c_1\ \Sigma\, d^2 \\ a &= \bar{S} - b\bar{T} - c_1\bar{D} \end{aligned}$$

Med tallene i tabel 1 som beregningsgrundlag får man følgende løsning:

$$\begin{aligned} a &= 9{,}1110 \\ b &= 0{,}2778 \\ c_1 &= 11{,}0556 \end{aligned}$$

Indsættes disse koefficienter i (1), kan man finde de enkelte kvartalers beregnede salg og sammenligne disse med de faktiske salg. Disse er opstillet i tabel 2 både for de otte analysekvartaler og for otte prognosekvartaler.

			Model 1		Model 2	
T	D	S	S-ber	Afvigelse	S-ber	Afvigelse
1	1	20	20,44	−0,44	20,00	0,0
2	0	10	9,67	0,33	9,86	0,14
3	0	10	9,94	0,06	10,07	−0,07
4	0	10	10,22	−0,22	10,29	−0,29
5	1	22	21,56	0,44	22,00	0,00
6	0	11	10,78	0,22	10,71	0,29
7	0	11	11,06	−0,06	10,93	0,07
8	0	11	11,33	−0,33	11,14	−0,14
9	1	24	22,67	1,33	24,00	0,00
10	0	12	11,89	0,11	11,57	0,43
11	0	12	12,16	−0,16	11,79	0,21
12	0	12	12,44	−0,44	12,00	0,00
13	1	26	23,78	2,22	26,00	0,00
14	0	13	13,00	0,00	12,43	0,57
15	0	13	13,28	−0,28	12,64	0,35
16	0	13	13,56	−0,56	12,86	0,14

Tabel 2. Forskelle mellem de faktiske og de beregnede salgstal.

Resultaterne i tabel 2 taler for sig selv. Det fremgår, at sæsonvariablen har fremkaldt nogenlunde overensstemmelse mellem de beregnede og de observerede salgstal.

Vi skal ikke stoppe her. Betragter man tabel 2, fremgår det, at afvigelserne i første-kvartalerne er voksende, hvilket er utilfredsstillende, og det antyder en mangel i modelspecifikationen. Årsagen er ganske oplagt. Modellen kan ikke afbilde afhængighed mellem trend og sæson. Nu er den afhængighed til stede i det anvendte dataeksempel, idet salget stiger med én enhed pr. år i kvartalerne 2, 3 og 4, medens det stiger med 2 enheder pr. år i kvartal 1. Det er altså nødvendigt at formulere en model, som kan afbilde denne afhængighed. Dette gøres meget enkelt ved at indføre en ekstra variabel H (H = TD) i model 1.

Værdierne af denne variabel fremgår af kolonne 4 i tabel 1. Den udvidede model har følgende udseende:

$$(2) \quad S = a + bT + c_1D + c_2H$$

Koefficienterne findes af det sædvanlige ligningssystem:

$$\begin{aligned} \Sigma\ st &= b\,\Sigma\, t^2 + c_1\,\Sigma\, td + c_2\,\Sigma\, th \\ \Sigma\ sd &= b\,\Sigma\, td + c_1\,\Sigma\, d^2 + c_2\,\Sigma\, dh \\ \Sigma\ sh &= b\,\Sigma\, th + c_1\,\Sigma\, dh + c_2\,\Sigma\, h^2 \end{aligned}$$

$$a = \overline{S} - b\overline{T} - c_1\overline{D} - c_2\overline{H}$$

Med tallene i tabel 1 som beregningsgrundlag fås følgende koefficienter:

$$\begin{aligned} a &= 9{,}4285 \\ b &= 0{,}2143 \\ c_1 &= 10{,}0714 \\ c_2 &= 0{,}2857 \end{aligned}$$

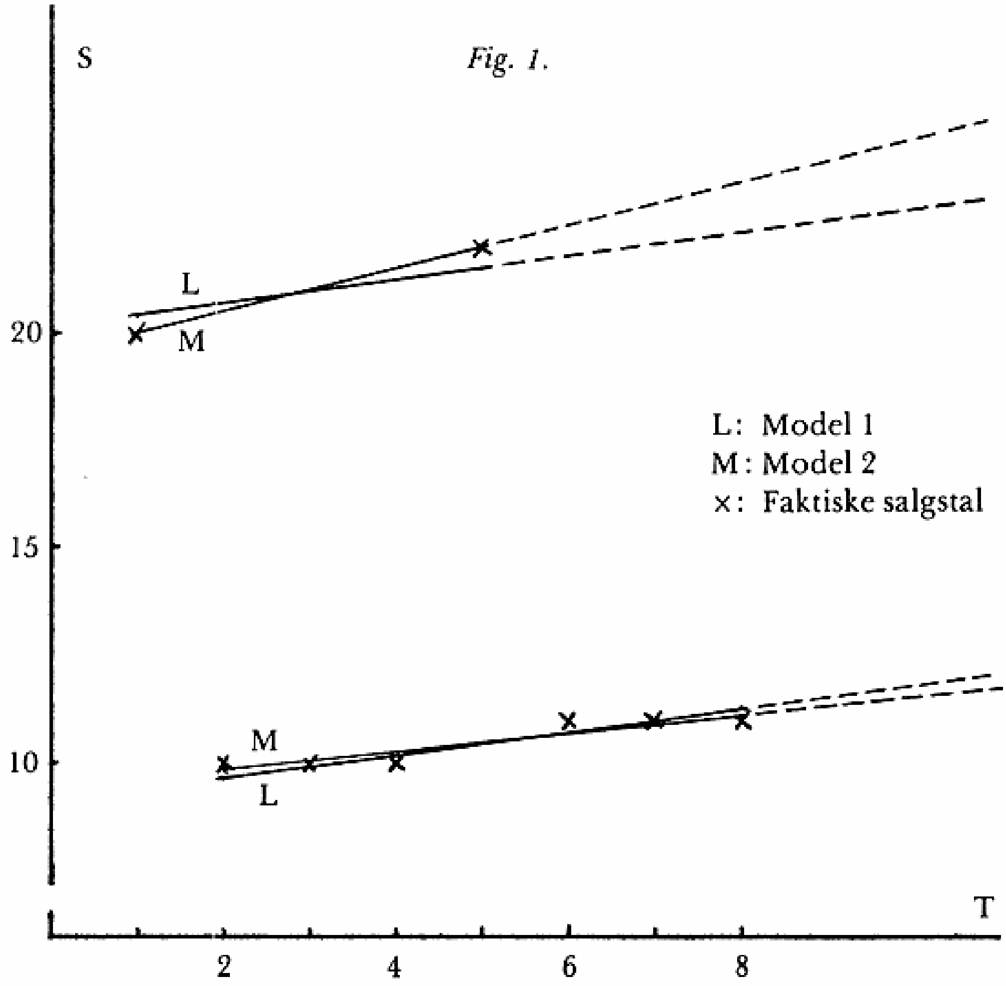
Indsættes disse resultater i (2), finder man det beregnede salg i kvartalerne 2, 3 og 4 – hvor D jo er lig nul – af:

$$S = 9{,}4285 + 0{,}2143T$$

I første-kvartalerne er D = 1, hvorfor salget beregnes af:

$$S = 19{,}4999 + 0{,}5000T$$

De beregnede salgstal efter model 2 er anført i tabel 2 sammen med afvigelserne fra dataeksemplets salgstal. Det ses, at modellen har ramt lige i prik i første-kvartalerne. I de øvrige kvartaler er afvigelserne af samme størrelsesorden som under model 1, men der kan konstateres en svag tendens mod underestimering i prognoseperioden. Alt i alt må man dog karakterisere resultatet som tilfredsstillende.
For illustrationens skyld er de beregnede salgstal efter model 1 og 2 indtegnet i fig. 1 sammen med de faktiske salgstal.


S
Fig. 1.
L
M
20
L: Model 1
M: Model 2
×: Faktiske salgstal
15
M
10
L
T
2
4
6
8

Afsluttende bemærkninger

Mindste kvadraters metode er naturligvis ikke et universelt svar på prognoseproblemer af enhver art. Men som vist i det foranstående er den meget anvendelig til analysering af den type prognoseproblemer, som Bent Vabø behandlede i den tidligere omtalte artikel. De vilkårlige resultater, som blev opnået der, taler ikke til ugunst for MKM, idet de udelukkende er et resultat af en mangelfuld modelspecifikation. Det er jo altid vigtigt, at man får samtlige betydningsfulde forklarende variable med i modellen, uanset om man behandler prognoseproblemer eller andre problemtyper. I de tilfælde, hvor man ikke a priori kender sæsonbevægelsernes struktur, vil man således starte med en model, der kan afbilde forskelle mellem alle delperioder. Viser det sig, at nogle delperioder er ens m.h.t. sæsonindflydelse – som i det anvendte eksempel – kan modellen reduceres tilsvarende.

Bent Vabø anbefaler – ganske vist på et forkert grundlag – at man beregner glidende gennemsnit (GG) fremfor at anvende MKM. Sammenlignes de to metoder, kan for det første bemærkes, at GG jo kun er et første skridt på vejen til bestemmelse af trenden. For det andet giver MKM – i den her benyttede udformning – detaljeret information om sæsonudsvingenes størrelser. Disse kan i det almindelige tilfælde testes med de sædvanlige statistiske testmetoder.

Det er rigtigt, at man mister frihedsgrader, når modellen udvides med sæsonvariable, men nøjagtigt det samme tab opstår ved beregning af GG. Er der således fire delperioder som i eksemplet, kan tre af observationerne ikke aktiveres i beregningen af GG.

En af årsagerne til at MKM er blevet så populær, er klart nok, at den er forholdsvis simpel at behandle matematisk. En anden tungtvejende grund er imidlertid, at den resulterer i maximum-likelihood estimater, d.v.s. estimater, som maximerer sandsynligheden for, at den hypotetiske population afkaster de observerede data (in casu tidsserier). De hermed forbundne egenskaber modsvarer ikke altid de ønskede egenskaber. Man vil især finde afvigelser indenfor det økonomiske område, idet opgaven her ofte består i at minimere et omkostningsbegreb, som ikke nødvendigvis er proportionalt med summen af de kvadrerede afvigelser mellem beregnede og observerede værdier. Idealet er selvfølgeligt, at der opstilles en tabsfunktion, som er specielt tilpasset det problem, der er under behandling. Dette fører ofte til uløselige eller vanskeligt løselige modelformuleringer. I stedet for at give helt op vælger man da en procedure, som giver en tilnærmet løsning, hvilken, i forbindelse med de her behandlede problemtyper, med fordel kan være mindste kvadraters metode.