Kristina Birch (Center for Statistik, CBS)
Jørgen Kai Olsen (Institut for Afsætningsøkonomi, CBS)
Tue Tjur (Center for Statistik, CBS)

Regressionsmodeller for Markedsandele

0. Introduktion.

I 122 uger (fra juni 1994 til november 1996) indsamlede markedsanalyseinstituttet Millward Brown Denmark et datasæt bestående af ugentlige totalsalg og gennemsnitspriser for de tre dominerende kaffemærker i Danmark, Gevalia (G), Merrild (M) og Karat (K). Desuden registrerede man det totale salg og en gennemsnitspris for alle andre mærker "Andre" (A). Data kan man finde på netadressen `www.mes.cbs.dk/~sttt/Coffee.txt.`

Formålet med dette symposieindlæg er at give et bud på hvordan man analyserer data af denne type, og — især — hvordan resultaterne kan fortolkes indenfor rammerne af almindelig mikroøkonomisk teori.

I afsnit 1 diskuteres nogle simple mikroøkonomiske modeller og deres egenskaber. Vores favorit blandt disse er modellen (5) som præsenteres i underafsnit 1.3. Denne model har, som alle de andre vi betragter, den egenskab at den inducerede model for markedsandele er en variant af "MCI–modellen" (the multiplicative competitive–interaction model, Cooper 1988, 1993). Men derudover har den, som model for de totale salgstal, nogle egenskaber der forekommer ganske realistiske.

Denne model og nogle af dens udvidelser kan tages som begrundelse for, at de relevante statistiske modeller er multiple regressionsmodeller med logaritmiske salgstal eller markedsandele som responser og logaritmiske priser som forklarende variable. I afsnit 2 diskuterer vi fortolkningen af disse modeller og nogle af de mange tests man kan foretage i dem.

1. Deterministiske modeller.

Lad $b = 1, \ldots, B$ (b for *brand*) betegne mærkerne for et givent "fmcg" (fast moving consumer good). Lad Q_{bt} betegne antallet af vareenheder som er solgt (f.eks. i Danmark) af mærke b i en kort periode (typisk en uge) t, og lad p_{bt} betegne gennemsnitsprisen for en enhed af mærke b i denne periode.

1.1. Den simple model.

Vores udgangspunkt er den matematisk meget simple — men desværre helt urealistiske — model, der beskriver salget som funktion af prisen ved Cobb–Douglas efterspørgselsfunktionen

$$Q_{bt} = \exp(\delta_t + \alpha_b + \beta_b \log(p_{bt})) \,. \qquad (1)$$

Her er $\beta_b = \frac{d \log Q_{bt}}{d \log p_{bt}}$ priselasticiteten, hvis fortolkning er at en lille relativ ændring af prisen fra p_b til $(1 + \Delta)p_b$ ville bevirke et fald af salget fra Q_{bt} til $(1 + \beta_b\Delta)Q_{bt}$. Det ligger altså i sagens natur, at β_b er negativ. Parameteren α_b, som man kunne kalde for "mærkets styrke", kan opfattes som et samlet mål for alle de tidsuafhængige (eller i det mindste langsomt varierende) faktorer, der påvirker salget ud over prisen, såsom kvalitet, udbredelse og forsyningssikkerhed, renommé, reklameindsats osv. Parameteren δ_t tager højde for de ændringer fra uge til uge, som virker på det samlede salg i hele kategorien og antages at fordele sig proportionalt på de enkelte mærker (specielle højtider og helligdage, fjernsynsprogrammer, vejret, årstiden osv.). Det er dette uundgåelige led, der er grunden til at vi omtaler vores modeller som modeller for markedsandele, selvom vi formulerer dem som modeller for totale salgstal. Da vi ikke (i denne artikel) har ambitioner om at forudsige den proportionale uge–til–uge effekt $\exp(\delta_t)$, er det i virkeligheden kun *forholdene* mellem salgene af de enkelte mærker vi forsøger at sige noget om.

Denne simple model er desværre ubrugelig til beskrivelse af et marked med mere eller mindre substituerbare mærker. For at forstå hvorfor, kan man nøjes med at se på tilfældet hvor alle mærker har samme elasticitet $\beta_b = \beta$. I dette tilfælde bliver det *samlede* salg af det pågældende forbrugsgode

$$Q_{1t} + \cdots + Q_{Bt} = \sum_{b=1}^{B} \exp(\delta_t + \alpha_b + \beta \log(p_b))$$

hvoraf umiddelbart følger, at en relativ tilvækst af *alle* mærkers priser med en faktor $1 + \Delta$ ville føre til et relativt fald af det totale salg med faktoren $1 + \beta\Delta$. Altså at det samlede salg i kategorien styres af samme priselasticitet som det enkelte mærkes salg. Det er naturligvis urealistisk. Når det gælder kaffe kan man jo prøve at forestille sig, at et enkelt mærke pludselig satte prisen op til det dobbelte, uden at de andre ændrede på prisen. Det ville givetvis føre til noget nær den totale ruin for den pågældende producent. Hvis derimod *alle* mærker fordoblede deres priser, ville ændringen af det samlede kaffeforbrug nok være mindre dramatisk. Faktisk skete der noget i den retning i 1994 og 1997, hvor en meget dårlig kaffehøst førte til prisstigninger på omkring 50%, uden at forbruget faldt nævneværdigt af den grund.

1.2. Krydselasticiteter.

Introduktionen af krydselasticiteter kan opfattes som et forsøg på at løse ovennævnte problem. For at tage højde for det fænomen, at et prisfald for et enkelt mærke ikke kun vil øge salget af det pågældende mærke, men også formindske salget af de andre mærker, udvides modellen til

$$Q_{bt} = \exp\left(\delta_t + \alpha_b + \beta_b \log(p_{bt}) + \sum_{k:k\neq b} \gamma_{bk} \log(p_{kt})\right) . \tag{2}$$

Fortolkningen af krydselasticiteterne $\gamma_{bk} = \frac{d \log Q_{bt}}{d \log p_{kt}}$, $b \neq k$, er åbenbart at hvis mærke k forøger sin pris fra p_k til $(1+\Delta)p_k$, så vil salget af mærke b ændre sig fra Q_{bt} til $(1+\gamma_{bk}\Delta)Q_{bt}$. Krydselasticiteterne forventes således at være positive.

Som model for de totale salg Q_{bt} er denne model mere realistisk end den simple model (1). Men ser man nøjere efter har den nogle ejendommelige egenskaber, der gør den uanvendelig til andet end lokal approksimation.

Ved differentiation af det totale salg m.h.t. $\log(p_{bt})$ fås

$$\frac{d(Q_{1t} + \cdots + Q_{Bt})}{d\log(p_{bt})} = \beta_b Q_{bt} + \sum_{k:k\neq b} \gamma_{kb} Q_{kt} \,.$$

Problemet er, at hvis vi forestiller os en fortsat stigning af prisen for mærket b, så vil Q_{bt} på et tidspunkt blive så lille at denne afledede bliver positiv (størrelserne γ_{kb} er jo positive, og salgene af de øvrige mærker er stigende). En moderat stigning af prisen på mærke b — som ikke modsvares af prisstigninger for de øvrige mærker — vil typisk (også ifølge modellen) føre til at salget af mærke b falder, og en del af dette salg overtages af konkurrenterne. Men så snart den afledte ovenfor bliver positiv, vil det samlede salg altså stige ifølge modellen, og det forekommer jo noget mærkværdigt.

Bemærk at modellen er overparametriseret i følgende forstand: Hvis vi for et givet mærke b ændrer β_b til $\beta_b + \kappa$, og samtidig ændrer alle krydselasticiteterne γ_{kb} til $\gamma_{kb} + \kappa$, så svarer dette blot til at multiplicere *alle* salgstal med faktoren $\exp(\kappa \log p_{bt})$, hvilket vi kan kompensere for ved at ændre den logaritmiske tidstrend δ_t til $\delta_t - \kappa \log p_{bt}$. Selve parametrene β_b og γ_{kb} er således ikke identificerbare, det er differenser som $\beta_b - \gamma_{kb}$ og $\gamma_{k'b} - \gamma_{k''b}$ der bestemmer markedsandelene. Formelt kan man fjerne denne overparametrisering ved at antage $\beta_b = 0$, således at der kun indgår krydselasticiteter i modellen. Men det gør ikke fortolkningen nemmere.

1.3. En mere realistisk model.

Markedsandelene

$$S_{bt} = \frac{Q_{bt}}{Q_{1t} + \cdots + Q_{Bt}}$$

er ifølge den simple model (1) givet ved

$$S_{bt} = \frac{\exp(\alpha_b + \beta_b \log(p_{bt}))}{\sum_{k=1}^{B} \exp(\alpha_k + \beta_k \log(p_{kt}))} \tag{4}$$

Dette er MCI–modellen, en særdeles nyttig og ofte anvendt model for markedsandele. Hvis man tager udgangspunkt i modellen (2), som også indeholder krydselasticiteter, får man en udvidet variant af MCI–modellen, som også undertiden benyttes i analysen af denne type data. Vores problem er, at da både (1) og (2) er urealistiske, giver de ikke nogen god begrundelse for anvendelsen af MCI–modellerne. Der findes imidlertid en model for de totale salg, som har langt mere realistiske egenskaber, og som fører til præcis den samme model (4) for markedsandelene, nemlig

$$Q_{bt} = \exp(\delta_t) \frac{\exp(\alpha_b + \beta_b \log(p_{bt}))}{\exp(\rho_t) + \sum_{k=1}^{B} \exp(\alpha_k + \beta_k \log(p_{kt}))} \tag{5}$$

Højre side ligner meget højre side i modellen (4) for markedsandele. Men der er to vigtige forskelle: Det additive led $\exp(\rho_t)$ i nævneren, og den multiplikative faktor $\exp(\delta_t)$ foran brøken. For at forstå denne model kan man tænke sig listen $\{1, \ldots, B\}$ af mærker udvidet med et "pseudomærke" som vi kalder mærke 0. At "købe mærke 0" betyder at man slet ikke køber. I denne fortolkning er

$$Q_{0t} = \exp(\delta_t) \frac{\exp(\rho_t)}{\exp(\rho_t) + \sum_{k=1}^{B} \exp(\alpha_k + \beta_k \log(p_{kt}))}$$

antallet af vareenheder (i vores eksempel 500g kaffeposer) som *ikke* blev købt i uge t (fordi kaffen generelt var for dyr). Det multiplikative led

$$\exp(\delta_t) = Q_{0t} + Q_{1t} + \cdots + Q_{Bt}$$

er den øvre grænse for antallet af enheder man kan sælge i den pågældende uge. Eller, om man vil, salget i en hypotetisk situation hvor alle priser er så lave, at de i realiteten ikke har nogen indflydelse på forbrugernes valg.

En udvidet fortolkning går ud på, at Q_{0t} inkluderer salget af mærker som vi ikke har oplysninger om. Leddet $\exp(\rho_t)$ i nævneren kan så fortolkes som en sum af bidrag af formen $\exp(\alpha_k + \beta_k \log(p_{kt}))$ for disse ikke–observerede mærker. Og man kan også bruge dette led til opsamling af mere diffuse alternativer. Fortolkningen af køb af mærke 0 som "ikke at købe" eller "købe et uobserveret mærke" er kun de to yderpunkter af en skala, som i tilfældet kaffe kunne omfatte alternativer som køb af pulverkaffe, te eller kakao.

Lad

$$S'_{bt} = \frac{Q_{bt}}{Q_{0t} + Q_{1t} + \ldots Q_{Bt}}$$

betegner de "generaliserede markedsandele" (hvor mærke 0 tæller med i nævneren). De direkte priselasticiter i modellen (5) ses da at være

$$e_{bb} = \frac{d\log(Q_{bt})}{d\log(p_{bt})} = \beta_b\,(1 - S'_{bt})$$

medens krydselasticiteterne er

$$e_{bk} = \frac{d\log(Q_{bt})}{d\log(p_{kt})} = -\beta_k S'_{kt}$$

Forudsat at alle β_b er negative, får disse størrelser således de fortegn de skal have, og krydselasticiteterne har desuden den meget ønskværdige egenskab at de er små for mærker med små markedsandele. Som følge heraf har modellen også den egenskab, som modellen (2) manglede, at en isoleret prisstigning for et enkelt mærke aldrig kan gøre det samlede salg større: Differentiation af det totale salg med hensyn til $\log(p_{bt})$ giver

$$\frac{d(Q_{1t} + \cdots + Q_{Bt})}{d\log(p_{bt})} = \exp(\delta_t)\beta_b S'_{bt} S'_{0t}$$

som altid er negativ, forudsat at β_b er negativ.

For at undersøge modellens egenskaber yderligere kan vi studere dens opførsel i en situation, hvor alle priser varierer fra periode til periode, på en sådan måde at forholdene mellem dem er konstante. Vi antager altså at

$$p_{bt} = \lambda_t p_b$$

hvor λ_t er den fælles proportionalitetsfaktor. Så er

$$S_{bt} = \frac{Q_{bt}}{Q_{1t} + \cdots + Q_{Bt}} = \frac{\exp(\alpha_b + \beta_b \log(\lambda_t p_b))}{\sum_{k=1}^{B} \exp(\alpha_k + \beta_k \log(\lambda_t p_k))}\,.$$

Vi bemærker først, at hvis alle parametrene β_b er ens, så forkorter faktoren $\exp(\beta_b \log(\lambda_t))$ ud, så markedsandelene bliver konstante. Hvis alle priser falder (proportionalt) til et meget lavt niveau vil det samlede salg nærme sig $\exp(\delta_t)$, og hvis priserne stiger proportionalt ud over alle grænser vil det samlede salg naturligvis falde til 0. Men forholdene mellem de forskellige mærkers salgstal vil være det samme hele vejen.

For fortolkningen af parametrene β_b er det måske mere interessant hvad der sker hvis de ikke er ens. Udtrykket for S_{bt} viser, at hvis β_b'erne er negative og λ_t går imod 0, så vil mærket med den højeste værdi af $|\beta_b|$ i grænsen overtage hele markedet. Omvendt, hvis alle priser vokser proportionalt til svimlende højder, så er det mærket med den laveste værdi af $|\beta_b|$ som i grænsen overtager hele markedet (altså relativt set, markedet er jo i mellemtiden næsten forsvundet).

Generelt kan man således sige, at et mærke med en lav værdi af $|\beta_b|$ er robust (hvad angår markedsandel) når priserne vokser proportionalt til et højt niveau, medens mærker med en høj værdi af $|\beta_b|$ er robuste når priserne aftager proportionalt til et meget lavt niveau. Det er nærliggende at karakterisere de første som discount–mærker, der overtager markedet når priserne er så høje at det gør ondt, medens de sidste måske er de mærker som alle ville vælge hvis de havde råd. Kort sagt, $|\beta_b|$ kan i nogen grad fortolkes som et mål for kvalitet. Men denne karakterisering er ikke helt udtømmende, for i princippet kan et mærke godt have en høj pris og en lav $|\beta_b|$–værdi, eller omvendt.

2. Statistiske modeller.

Lad $Q_{\mathrm{G}t}$, $Q_{\mathrm{M}t}$, $Q_{\mathrm{K}t}$ og $Q_{\mathrm{A}t}$ betegne salgene af de fire mærker (idet betegnelsen "mærke" altså også bruges om restgruppen "Andet") i uge $t = 1, 2, \ldots, 122$. Med $l_{\mathrm{G}t}$, $l_{\mathrm{M}t}$, $l_{\mathrm{K}t}$ og $l_{\mathrm{A}t}$ betegner vi (for kortheds skyld) logaritmerne til priserne.

2.1. Multiple regressionsmodeller for log(salg).

Ved at tage logaritmen på begge sider af (5) får vi

$$\log Q_{bt} = \delta_t' + \alpha_b + \beta_b l_{bt}$$

hvor de nye parametre δ_t' er defineret som de gamle δ_t minus logaritmen til nævneren på højre side. At vi på denne måde introducerer en ny "tidstrend" δ_t' med en højst kompliceret fortolkning er uden betydning, da vi alligevel opfatter tidsvariationen som 122 "støjparametre", hvis eneste formål er at korrigere for uge–til–uge variationen. Det væsentlige er, at nævneren på højre side af (5) ikke afhænger af b.

Det ligger lige for at forsyne ligningen ovenfor med et fejlled på højre side og opfatte den som specifikationen af en multipel regressionsmodel. Så vi antager nu at $\log Q_{bt}$ er stokastisk, givet ved

$$\log Q_{bt} = \delta_t' + \alpha_b + \beta_b l_{bt} + \varepsilon_{bt} \tag{6}$$

hvor de 4×122 stokastiske variable ε_{bt} er uafhængige, normalfordelte med middelværdi 0 og samme varians σ^2. Altså en additiv tosidet (4×122)

variansanalysemodel med en ekstra kovariat, hvis koefficient afhænger af rækkenummeret.

Varianshomogeniteten er selvfølgelig noget der bør undersøges. Den viser sig at holde udmærket i dette datasæt, ligesom alle andre antagelser ser ud til at være opfyldt for de modeller vi betragter i det følgende. Men modelkontrollen er der ikke plads til at gå i detaljer med her.

Før vi kaster os ud i en analyse af denne model, er det imidlertid værd at bemærke, at den har en naturlig udvidelse, som bør estimeres først. Nemlig modellen

$$\log Q_{bt} = \delta_t' + \alpha_b + \beta_b l_{bt} + \sum_{k:k\neq b} \gamma_{bk} l_{kt} + \varepsilon_{bt} \,. \tag{7}$$

Fra et statistisk synspunkt forekommer denne model meget naturlig, idet den jo blot udvider model (6) med nogle forklarende variable, som sagtens kan tænkes at være relevante. Økonomisk er den lidt vanskelig at fortolke, ud over at den helt svarer til hvad man ville få ved at introducere "krydselasticitetsled" i (5). Bemærk at den har en lidt usædvanlig kovariatstruktur, idet de i alt fire kovariater af længde 4×122 har værdier der stammer fra de fire "log(pris)–vektorer" af længde 122. Hver værdi genbruges altså fire gange.

Uanset fortolkningen gælder jo i hvert fald, at hvis vi ikke kan få godkendt modelreduktionen fra (7) til (6), så er der noget galt med (6).

Estimaterne for β'er og γ'er i modellen (7) bliver (idet overparametriseringen er den samme som for krydselasticitetsmodellen i afsnit 1.2, så vi kan nøjes med at angive estimater for differenserne $\beta_b - \gamma_{kb}$)

Parameter	Estimate	Std.dev.	T	P
$\beta_{\mathrm{G}} - \gamma_{\mathrm{MG}}$	-4.862	0.6116	-7.951	0.000000
$\beta_{\mathrm{G}} - \gamma_{\mathrm{KG}}$	-4.603	0.6116	-7.527	0.000000
$\beta_{\mathrm{G}} - \gamma_{\mathrm{AG}}$	-3.316	0.6116	-5.423	0.000000
$\beta_{\mathrm{M}} - \gamma_{\mathrm{GM}}$	-5.764	0.7016	-8.216	0.000000
$\beta_{\mathrm{M}} - \gamma_{\mathrm{KM}}$	-6.148	0.7016	-8.764	0.000000
$\beta_{\mathrm{M}} - \gamma_{\mathrm{AM}}$	-4.500	0.7016	-6.414	0.000000
$\beta_{\mathrm{K}} - \gamma_{\mathrm{GK}}$	-3.705	0.5949	-6.228	0.000000
$\beta_{\mathrm{K}} - \gamma_{\mathrm{MK}}$	-3.808	0.5949	-6.402	0.000000
$\beta_{\mathrm{K}} - \gamma_{\mathrm{AK}}$	-3.418	0.5949	-5.747	0.000000
$\beta_{\mathrm{A}} - \gamma_{\mathrm{GA}}$	-0.737	0.4923	-1.497	0.135302
$\beta_{\mathrm{A}} - \gamma_{\mathrm{MA}}$	-1.803	0.4923	-3.663	0.000288
$\beta_{\mathrm{A}} - \gamma_{\mathrm{KA}}$	-0.617	0.4923	-1.254	0.210851

De tests der er foretaget i denne tabel for hypoteser af formen $\gamma_{kb} = \beta_b$ er ret irrelevante. Den hypotese vi interesserer os for (modellen (6)) svarer til, at parametrene er ens tre og tre (de tre første ens, de næste tre ens osv.). Men den bliver forkastet i dette tilfælde ((F(8,351)=5.305, P=0.000003).

Ser man nærmere på estimaterne, er det tydeligt at den mest udtalte afvigelse fra hypotesen går ud på at parametrene $\beta_b - \gamma_{kb}$, b fast, er mindre i absolut værdi for $k = \mathrm{A}$ end når k er et egentligt kaffemærke. For eksempel har (som man ser af tabellens tre første linier) prisen på Gevalia større indflydelse på Merrilds og Karats markedsandele end på markedsandelen for restgruppen "Andet". Det er måske ikke så forbavsende, når man tænker på at "Andet" jo står for en ophobning af adskillige mindre kaffemærker, hvor i hvert fald dem der er knyttet til bestemte butikskæder (Blå Cirkel, Irma kaffe osv.) må antages at have ret loyale kunder. Vi kan altså ikke få modellen (6) til at passe, hvis restgruppen skal have samme status som de egentlige mærker.

Som nævnt i afsnit 1.3 kræver modellen (5) (og dermed (6)) imidlertid ikke, at alle mærker på markedet er med. Statistisk betyder dette, at vi kan udelade data for restgruppen, idet gennemsnitsprisen for "Andet" så må opfattes som en af de mange ukendte variationskilder der er skjult i modellens restled. Vi anvender derfor modellen (7) på det datasæt af længde 3×122, som består af oplysningerne for de tre egentlige mærker. Resultatet bliver nu

Parameter	Estimate	Std.dev.	T	P
$\beta_{\mathrm{G}} - \gamma_{\mathrm{MG}}$	-4.864	0.6503	-7.480	0.000000
$\beta_{\mathrm{G}} - \gamma_{\mathrm{KG}}$	-4.603	0.6503	-7.078	0.000000
$\beta_{\mathrm{M}} - \gamma_{\mathrm{GM}}$	-5.139	0.6800	-7.558	0.000000
$\beta_{\mathrm{M}} - \gamma_{\mathrm{KM}}$	-5.453	0.6800	-8.020	0.000000
$\beta_{\mathrm{K}} - \gamma_{\mathrm{GK}}$	-3.730	0.6226	-5.991	0.000000
$\beta_{\mathrm{K}} - \gamma_{\mathrm{MK}}$	-4.061	0.6226	-6.523	0.000000

Her ser det ud som om estimaterne med god approksimation er parvis ens, og et F–test bekræfter dette: F(3,236) = 0.285, svarende til en P–værdi på 0.84. Vi får altså godkendt hypotesen

$$\gamma_{\mathrm{MG}} = \gamma_{\mathrm{KG}}, \quad \gamma_{\mathrm{GM}} = \gamma_{\mathrm{KM}} \text{ og } \gamma_{\mathrm{GK}} = \gamma_{\mathrm{MK}}$$

svarende til model (6) (idet vi herefter kan sætte γ–parametrene til 0).

I denne model kunne vi nu prøve at tilføje et led af formen $\gamma_b l_{\mathrm{A}t}$, for at undersøge om gennemsnitsprisen i restgruppen påvirker forholdene mellem de tre store mærkers salgstal. Men da effekten af denne kovariat kun er svagt signifikant (P=0.035), tager vi ikke dette led med i den endelige model. Den effekt, som restgruppens gennemsnitspris uden tvivl har, fordeler sig åbenbart næsten proportionalt på de tre mærker.

Estimaterne i model (6) for det reducerede datasæt bliver

Parameter	Estimate	Std.dev.	T	P
β_{G}	-5.013	0.4485	-11.176	0.000000
β_{M}	-5.019	0.4695	-10.691	0.000000
β_{K}	-3.880	0.4144	-9.364	0.000000

Her kunne man videre interessere sig for, om hypotesen $\beta_{\mathrm{G}} = \beta_{\mathrm{M}} = \beta_{\mathrm{K}}$, svarende til "samme elasticitetsformel for de tre mærker", kan godkendes. Det kan den ikke, viser det sig (P=0.0017). Derimod kan vi godt få godkendt at $\beta_{\mathrm{G}} = \beta_{\mathrm{M}}$. Karat er åbenbart lidt mindre priselastisk end de to andre mærker. Det betyder, som nævnt sidst i afsnit 1.3, at Karats salgstal procentvis falder mindre end salgstallene for de to andre mærker, i en situation hvor alle priser vokser proportionalt. Det har muligvis noget at gøre med Karats særlige rolle som lokalt mærke i det nordlige Jylland (produceret i Aalborg). Hvis Karat har en forholdsvis loyal kundeskare, kan man forestille sig at Gevalia og Merrild er mere sårbare over for konkurrence fra discount–mærker i restgruppen.

2.2. Tilføjelse af andre kovariater.

Vores forsøg på at tilføje kovariaten $l_{\mathrm{A}t}$ til modellen ovenfor er bare ét eksempel på, hvordan modellen (6) kan udvides med forklarende variable, som kan tænkes at have indflydelse på markedsandelene. Et oplagt forslag ville være at inddrage et eller andet mål for reklametrykket for de enkelte mærker, for eksempel i form af producenternes udgifter til reklame eller logaritmen til disse. Desværre har vi ikke den slags oplysninger. Men vi kan illustrere idéen ved tilføjelse af tre variable som vi har, nemlig de "laggede" log(priser). Vi opstiller altså nu modellen

$$\log Q_{bt} = \delta_t' + \alpha_b + \beta_b \log(p_{bt}) + \eta_b \log(p_{b,t-1}) + \varepsilon_{bt} \qquad (8)\,.$$

Estimaterne i denne model er

Parameter	Estimate	Std.dev.	T	P
β_{G}	-5.188	0.5508	-9.419	0.000000
β_{M}	-5.802	0.5793	-10.016	0.000000
β_{K}	-4.020	0.4980	-8.072	0.000000
η_{G}	1.931	0.5565	3.470	0.000619
η_{M}	2.747	0.5704	4.816	0.000003
η_{K}	1.832	0.4946	3.704	0.000265

Umiddelbart virker det måske lidt overraskende, at de laggede priser har signifikant indflydelse på forbrugernes mærkevalg. Man skulle jo ikke tro, at priserne fra sidste uge er afgørende for hvilket mærke man vælger. Men tænker man lidt efter, er der en ret oplagt forklaring på at koefficienterne η_b er positive. Det har rimeligvis noget at gøre med, at når ens yndlingsmærke pludselig er på tilbud, så køber man det, også selvom man egentlig havde tænkt sig at vente til næste uge, og man køber måske endda fire poser i stedet for en. Hvis vi skriver modellen på formen

$$\log Q_{bt} = \delta_t + \alpha_b + (\beta_b + \eta_b)\log(p_{bt}) + \eta_b \log\left(\frac{p_{b,t-1}}{p_{bt}}\right) + \varepsilon_{bt}$$

bliver det lidt mere gennemskueligt hvad der foregår. Parametrene $\beta_b + \eta_b$ er dem der — via egentlige (langtids–) elasticiteter — bestemmer hvordan de tre mærker deler markedet i en længere periode med konstante priser ($\log \frac{p_{b,t-1}}{p_{bt}} = 0$). Parameteren η_b bestemmer korttidsreaktionen på en prisændring. At η_b'erne er positive betyder, at det mersalg, som udløses af et prisfald, er større på kort end på langt sigt.

De egentlige (langtids–) elasticiteter vil således typisk være væsentligt mindre i absolut værdi, end man får indtryk af ved estimation af modellen (6). Det, vi er faldet over her, er derfor ikke bare en lille pudsighed, der illustrerer hvordan model (6) kan udvides. Det er en korrektion til modellen, som er helt nødvendig, hvis den skal kunne bruges som et nogenlunde præcist redskab i forbindelse med valg af prisstrategi.

Vi har naturligvis også prøvet at tilføje de dobbeltlaggede logaritmiske priser (altså log(pris) fra to uger før) som kovariater, men de viste sig at være insignifikante.

Referencer.

Birch, K., Olsen, J.K. and Tjur, T. (2005)
Regression Models for Market-Shares.
Preprint 1/2005, Center for Statistics, Copenhagen Business School.
`http://staff.cbs.dk/tuetjur/05-1.pdf`

Cooper, Lee G. and Nakanishi, Masao (1988).
Market–Share Analysis.
Kluwer Academic Publishers.

Cooper, Lee G. (1993).
Market–Share Models.
in *Handbook in Operations Research and Management Science* **vol. 5** (Marketing, ed. J. Eliashberg and G. L. Lilien), 259–314.

D. McFadden (1986)
The choice theory approach to market research
Marketing Science **Vol. 5** pp 274-297.