Statistikgruppen
Handelshøjskolen i København

MPAS
Tue Tjur
September 2002

## Notat 1:

### Modeller i forbindelse med projekt 1.

Der er (mindst) tre spørgsmål som kan analyseres i forbindelse med projektets problemstilling og de udleverede data, nemlig

1. Hvordan afhænger det faktiske salg af de styringsparametre (pris, reklame, særplacering) som vi og konkurrenten fastsætter?

2. Givet en model (inklusive nogle parameterestimater vi tør tro på) der besvarer spørgsmål 1, og givet de seks styringsparametre **samt** konkurrentens pris for en bestemt periode, hvordan fastsætter vi vores egen pris således at dækningsbidrag 2 maksimeres?

3. Selv om vi kan besvare spørgsmål 2 kan vi ikke direkte bruge det til noget, for vi kender ikke konkurrentens pris før det er for sent. Desuden er det ikke altid en god ide kun at styre efter maksimering af dækningsbidrag 2, fordi det kan føre til en hård konkurrencesituation som alle parter vil tabe på. En fornuftig strategi må inddrage mere langsigtede overvejelser, som også tager højde for hvad konkurrenterne kan finde på at gøre i fremtiden. Er det muligt at forudsige konkurrenternes pris? Og hvis det er muligt, hvordan skal vi så sætte vores egen pris, så den maksimerer dækningsbidrag 2 over en længere periode?

I dette notat beskæftiger vi os kun med spørgsmålene 1 og 2. Det tredje må vi se på senere.

### Notation.

For ikke at gøre formlerne for lange bruger vi følgende korte variabelbetegnelser:

| Her | I opgaveteksten |
|---|---|
| $i$ | `PERIODE` |
| $p$ | `PRIS` |
| $p_k$ | `K_PRIS` |
| $r$ | `REKL` |
| $r_k$ | `K_REKL` |
| $t$ | `PLAC` |
| $t_k$ | `K_PLAC` |
| $s$ | `SALG` |
| $s_k$ | `TOTSALG` − `SALG` |

Vi forudsætter `PLAC` og `K_PLAC` omregnet til kr. og øre (ved multiplikation med 900, som beskrevet nederst i datafilen).

**En model for markedsandelen.**

Det totale salg $s + s_k$ regner vi i første omgang for givet. Der kan være brug for senere at opstille modeller, der beskriver det totale salg som funktion af tid, sæson og (muligvis) en passende gennemsnitlig markedspris, men vi forventer ikke at det totale salg over kortere perioder vil være væsentligt påvirket af priser og reklameindsatser. Derimod vil vi nok forvente, at vores markedsandel er stærkt afhængig af forholdet mellem vores egen pris og konkurrentens pris, og formodentlig også af forholdene mellem de beløb der bruges til reklame og særplacering. En model, der tager disse overvejelser i betragtning, og som ud fra almindelig økonomisk teori er nærmest kanonisk, går ud på at vi har sammenhængen

$$\log \frac{s}{s_k} = \alpha + \beta \log \frac{p}{p_k} + \gamma \log \frac{r}{r_k} + \delta \log \frac{t}{t_k}$$

Som modellen står her er den rent deterministisk. I praksis kan vi naturligvis aldrig opnå en sådan eksakt sammenhæng, og lige straks laver vi den om til en statistisk model. Men først lidt om fortolkningen. Det er et væsentligt træk ved modellen, at det er *forholdet* mellem vores og konkurrentens salg, der er bestemt ved *forholdet* mellem vores og konkurrentens pris, reklame og udgifter til særplacering. Det ligger således i modellen, at hvis f.eks. både vi og konkurrenten nedskærer reklamebudgettet med 20% eller fordobler prisen, så bevirker det ingen ændringer i forholdet mellem salgstallene. Et andet væsentligt træk ved modellen — som faktisk ville være obligatorisk hvis der kun var én konkurrent i stedet for tre — er, at hvis vi bytter om på os selv og konkurrenten, svarende til at vende bunden i vejret på de fire brøker i formlen, så får vi præcis samme model, bare set fra konkurrentens synspunkt. Parametrene $\beta$, $\gamma$ og $\delta$ bliver de samme som før, medens $\alpha$ skifter fortegn. Konstantleddet $\alpha$ må i denne sammenhæng fortolkes som en præference for vores ($\alpha > 0$) eller konkurrentens ($\alpha < 0$) vare, som ikke lader sig påvirke af priser eller reklameindsats (det kunne f.eks. være forskelle i kvalitet eller renommé).

I relation til den økonomiske litteratur vil formlen muligvis forekomme mere bekendt hvis vi skriver den på formen

$$\frac{s}{s_k} = \text{konst.} \times \left(\frac{p}{p_k}\right)^{\beta} \times \left(\frac{r}{r_k}\right)^{\gamma} \times \left(\frac{t}{t_k}\right)^{\delta}.$$

hvor konstanten på højre side er $e^{\alpha}$. Men hvis vi skal lave den om til en statistisk model forekommer det mere naturligt at bevare den på logaritmisk form og tilføje et fejlled på højre side. Dvs. at opfatte vores 24 værdier af $\log \frac{s}{s_k}$ som uafhængige observationer af normalfordelte stokastiske variable med samme varians og middelværdier givet ved højre side af den oprindelige formel. Det fører jo til en simpel multipel regressionsmodel.

En udvidelse af denne model, som kan være relevant fordi den kan hjælpe til at kompencere for at "konkurrenten" i vores datasæt ikke er én konkurrent men et gennemsnit af tre, kunne gå ud på at splitte de tre brøker op i tæller og nævner hver for sig og forøge antallet af parametre tilsvarende. Altså (idet vi nu skriver $\mathrm{E}\frac{s}{s_k}$ i stedet for $\frac{s}{s_k}$, svarende til at vi er gået over til en statistisk model),

$$\mathrm{E}\log\frac{s}{s_k} = \alpha + \beta\log p + \beta_k\log p_k + \gamma\log r + \gamma_k\log r_k + \delta\log t + \delta_k\log t_k$$

altså en multipel regressionsmodel med seks forklarende variable. Her svarer den oprindelige model så til at de tre hypoteser $\beta = -\beta_k$, $\gamma = -\gamma_k$ og $\delta = -\delta_k$ er opfyldt (hvilket man eventuelt kan teste).

Angående fortolkningen af parametrene: Det er klart at vi forventer en negativ værdi af $\beta$ og positive værdier af $\gamma$ og $\delta$, fordi salget naturligvis vil aftage når prisen vokser (vi ser bort fra en evt. "snobberi–effekt") og vokse med markedsføringsindsatsen (vi ser bort fra negativ reklameeffekt). Som vi senere skal se er der nogle yderligere bånd på parametrene som skal være opfyldt, hvis modellen skal give mening fra et økonomisk synspunkt (nemlig $\beta < -1$, $\gamma + \delta < 1$).

### Optimering af dækningsbidrag 2 under en simplificerende antagelse.

I det følgende antager vi at $\mathrm{E}\log\frac{s}{s_k}$ kan beskrives ved en model som ovenfor, og desuden lader vi som om vi kender de syv parametre $\alpha, \ldots, \delta_k$. I praksis skal det selvfølgelig forstås sådan at vi har estimeret dem, og (i mangel af bedre) lader som om de estimerede værdier er sande.

Ifølge modellen er (idet vi regner deterministisk, dvs. ignorerer fejlleddet)

$$s = s_k \exp(\ldots)$$

hvor " ... " her og i det følgende står for udtrykket

$$\alpha + \beta\log p + \beta_k\log p_k + \gamma\log r + \gamma_k\log r_k + \delta\log t + \delta_k\log t_k$$

— altså middelværdien i den statistiske model.

Lad $p_0$ betegne den pris vi selv betaler for 1 stk. af varen (inklusive andre variable omkostninger, dvs. omkostninger knyttet til den enkelte vareenhed, såsom distribution, mærkning etc.). Det er denne størrelse, som ifølge de oplysninger vi har fået kan sættes til 15 kr. Så er dækningsbidrag 2 — det der bliver tilbage når de variable udgifter og markedsføringsudgifterne er betalt —

$$\mathrm{DB2} = (p - p_0)s - r - t = (p - p_0)s_k\exp(\ldots) - r - t.$$

Lad os nu, til en begyndelse, gøre den antagelse at konkurrentens salgstal $s_k$ ikke afhænger af vores valg af styringsparametre. Det er en antagelse, som naturligvis ikke holder eksakt, for hvis vi forøger vores eget salgstal ved at sætte prisen lavt og bruge mange penge på reklamer, må konkurrentens salgstal nødvendigvis falde. Men hvis vores markedsandel er meget lille vil antagelsen være en acceptabel approksimation.

Ser vi nu på udtrykket for DB2 som funktion af $p$, idet $r$ og $t$ holdes fast, kan man let overbevise sig om at hvis $\beta$ er større end $-1$, så vil DB2 konvergere mod $+\infty$ for $p \to +\infty$. Hvis derimod $\beta < -1$ vil udtrykket som funktion af $p$ antage et maksimum et eller andet sted mellem $p_0$ og $+\infty$. Den optimale pris finder vi ved at sætte differentialkvotienten lig med 0. Vi får

$$\frac{d}{dp}\mathrm{DB2} = s_k \exp(\dots) + (p - p_0)s_k \frac{\beta}{p} \exp(\dots),$$

som er 0 for

$$1 + \frac{p - p_0}{p}\beta = 0$$

eller

$$\beta = -\frac{p}{p - p_0} \quad (< -1, \text{ OBS})$$

altså for

$$p = \frac{\beta}{1 + \beta} p_0$$

Denne ligning har en fortolkning i forbindelse med almindelig økonomisk teori. Her vil man normalt skrive den på følgende form:

$$\text{Optimal pris} = \frac{e_P}{1 + e_P} \times \text{Grænseomkostninger.}$$

"Grænseomkostningerne" ($p_0$) hedder sådan, fordi det er det det koster at producere eller købe 1 stk. vare mere, og $e_P = \beta$ er priselasticiteten (som i vores tilfælde ikke engang afhænger af prisen $P$, fordi vores model antager "isoelasticitet").

Under disse antagelser er det således muligt at finde et eksplicit udtryk for den optimale pris — endda et udtryk, som ikke afhænger af $r$ og $t$. Desværre går det ikke helt så let for de to andre styringsvariable $r$ og $t$. Ser vi på DB2 som funktion af (for eksempel) $r$, kan vi for det første bemærke at for $\gamma > 1$ vil DB2 konvergere mod $+\infty$ for $r \to +\infty$. I grænsetilfældet $\gamma = 1$ er DB2 som funktion af $r$ beskrevet ved en linie, hvilket heller ikke forekommer realistisk. For $\gamma < 1$ har funktionen et maksimum mellem 0 of $+\infty$. Differentiation af DB2 m.h.t. $r$ giver

$$\frac{d}{dr}\mathrm{DB2} = (p - p_0)s_k \frac{\gamma}{r} \exp(\dots) - 1$$

som er lig med 0 for

$$r = \gamma(p - p_0)s_k \exp(\dots) = \gamma(p - p_0)s$$

Denne ligning betyder, at parameteren $\gamma$ har en meget konkret fortolkning som den andel af overskuddet før markedsføringsudgifter, DB1=$(p - p_0)s_k$, som vi skal bruge til reklamer for at optimere DB2. Men da $r$ også indgår i exp$(\dots)$ på højre side kan vi ikke løse denne ligning explicit.

På samme måde ses at der for den optimal løsning må gælde

$$t = \delta(p - p_0)s_k \exp(\dots) = \delta(p - p_0)s$$

dvs. $\delta$ er den andel af DB1= $(p - p_0)s_k$ vi skal bruge til særplacering af varen for at få en optimal løsning. Heraf følger (ved addition af de to ligninger) at

$$r + t = (\gamma + \delta)(p - p_0)s_k.$$

Eftersom vi ikke kan bruge flere penge end vi har, får vi her den yderligere betingelse for eksistens af et pænt maksimum for DB2 at

$$\gamma + \delta < 1.$$

Vi kan så kun håbe på at parameterestimaterne fra modellen opfylder dette krav.

En simpel iterativ metode, som rimeligvis konvergerer mod en løsning, går ud på følgende. Udregn først den optimale pris $p$, og vælg passende startværdier for $r$ og $t$. Udregn det forventede salgstal $s$ med disse værdier indsat. Udregn så nye værdier af $r$ og $s$ som

$r = \gamma(p - p_0)s$ og
$t = \delta(p - p_0)s$.

Herefter udregnes den nye værdi af $s$ på basis af disse værdier, og således fortsættes. Det er klart, at hvis denne simple algoritme konvergerer, så ender vi med en løsning. Jeg gætter på at den konvergerer, forudsat at $\beta < -1$ og $\gamma$ og $\delta$ er positive med $\gamma + \delta < 1$.

**Optimering af dækningsbidrag 2 under mere realistiske antagelser.**

En antagelse som er noget mere realistisk end antagelsen om, at $s_k$ er uafhængig af styringsparametrene $p$, $r$ og $t$, går ud på at det totale salg $s + s_k$ er uafhængigt af disse parametre. Dette er en antagelse om at markedet som helhed er uelastisk, hvilket virker rimeligt hvis det drejer sig om en dagligvare som folk køber når de har brug for den, uanset prisen (inden for rimelige grænser). Antagelsen indebærer, at det kun er valget af mærke, men ikke det kvantum man køber, der påvirkes af pris, reklame og særplacering. Det kan ganske vist tænkes at særlige slagtilbud får folk til at købe "tre til to pakkers pris" eller lignende, men det opvejes så af at det varer længere tid før de køber noget igen.

Under denne antagelse skal vi åbenbart udtrykke vores salg som

$$s = (s + s_k)\frac{\exp(\dots)}{1 + \exp(\dots)}$$

og får så udtrykket

$$\mathrm{DB2} = (p - p_0)s - r - t = (p - p_0)(s + s_k)\frac{\exp(\dots)}{1 + \exp(\dots)} - r - t$$

for den størrelse vi skal optimere. Differentiation m.h.t. $p$, $r$ og $t$ giver

$$\frac{d}{dp}\mathrm{DB2} = (s + s_k)\frac{\exp(\dots)}{1 + \exp(\dots)} + (p - p_0)(s + s_k)\frac{\beta}{p}\frac{\exp(\dots)}{(1 + \exp(\dots))^2}$$

$$\frac{d}{dr}\mathrm{DB2} = (p - p_0)(s + s_k)\frac{\gamma}{r}\frac{\exp(\dots)}{(1 + \exp(\dots))^2} - 1$$

$$\frac{d}{dt}\mathrm{DB2} = (p - p_0)(s + s_k)\frac{\delta}{t}\frac{\exp(\dots)}{(1 + \exp(\dots))^2} - 1$$

Ligningerne som sætter disse differentialkvotienter lig med 0 kan nemt omskrives til

$$p = \frac{\beta/(1 + \exp(\dots))}{1 + \beta/(1 + \exp(\dots))}p_0$$

$$r = \frac{\gamma(p - p_0)s}{1 + \exp(\dots)}$$

$$t = \frac{\delta(p - p_0)s}{1 + \exp(\dots)}$$

Bemærk at formlerne ser ud ganske som i det simplere tilfælde hvor $s_k$ blev opfattet som konstant, blot er $\beta$, $\gamma$ og $\delta$ erstattet med $\beta/(1 + \exp(\dots))$, $\gamma/(1 + \exp(\dots))$ og $\delta/(1 + \exp(\dots))$.

Ligningerne kan ikke løses explicit, fordi $p$, $r$ og $s$ jo også indgår på højre side både i $s$ og i $\exp(\dots)$. Men en simpel iterativ metode til løsning af dem går ud på følgende. Udregn først $\exp(\dots)$ og $s$ for passende startværdier af $p$, $r$ og $t$. Indsæt disse værdier på højre side af lighedstegnene i tre ovenstående formler, og få herved ny værdier af $p$, $r$ og $t$ — som igen tages som startværdier osv. osv. Denne algoritme vil rimeligvis konvergere for de værdier af $\beta$, $\gamma$ og $\delta$ hvor der findes en optimal løsning.

### Og hvad skal vi så bruge det til?

Vi skal i hvert fald næppe opfatte det bogstaveligt som en metode til at finde den optimale strategi. Man kan sige at den løsning vi har fundet er optimal under forudsætning af at konkurrenten er både blind og døv. Og det er han/hun/de jo nok ikke. Men kendskabet til, hvordan denne løsning ser ud, giver et fingerpeg om hvilken retning man skal gå i for at forøge dækningsbidrag 2. Hvor langt man vil gå i den retning skal så vurderes i forhold til hvad man tror konkurrenterne vil gøre. Man kunne, til en begyndelse, forsøge at svare på spørgsmål i stil med det følgende:

Hvis vi, gennem hele perioden, havde ændret vore styringsparametre ved at flytte dem (f.eks.) 1/10 af vejen fra de faktiske hen imod de "optimale" værdier — hvor optimalitet nødvendigvis må udregnes på basis af passende forudsagte værdier af konkurrentens styringsparametre og det totale salg — hvilke ændringer ville det have givet anledning til, især når det gælder det samlede dækningsbidrag 2 gennem hele perioden, forudsat at denne lille ændring ikke førte til ændringer fra konkurrentens side?

Bemærk at der alene i denne simple idé ligger et par statistiske problemer. Vi får brug for modeller, der forudsiger konkurrentens valg af styringsparametre og det totale salg bedst muligt ud fra tallene fra de foregående perioder.

Næste skridt kunne så være at gennemføre de samme udregninger, hvor de indsatte værdier af konkurrentens styringsparametre ikke længere er de faktisk observerede, men nogle som er genererede ud fra vores forudsigelsesmodel, evt. med noget "støj" lagt oveni.