

Nogle muligheder i scanner data

Resumé

I artiklen gives en diskussion af en række af de muligheder for effektivisering af marketingaktiviteter, der er til stede for såvel mærkevarereudbydere som detaillister, ved udnyttelse af information fra scanner data.

Indledning

Datafangst med laserbaseret scanning teknologi ved salg af varer i et supermarked blev introduceret i USA i 1974. Hurtigt blev fordelene ved scanning kendt og inden for de følgende 10 år havde næsten alle amerikanske supermarkeder installeret scanning udstyr. I mellemtiden havde man også i Vesteuropa startet indførelsen af scanning af varer, og den seneste opgørelse for Danmark viser, at 94% af fødevarereomsætningen bliver scannet ved salg (Hawkes&Smith,1999).

De umiddelbare argumenter for indførelse af scannerudstyr var, at det gav forbedrede muligheder for lagerkontrol og -styring og ville være omkostningsbesparende, da det giver højere produktivitet ved kassen, og man kan undlade prismærkning af hvert eneste emne i butikken. Bestilling af nye varer kan automatiseres endog i visse tilfælde ved direkte opkobling til leverandører.

Der er imidlertid også en lang række

andre muligheder for detailhandlere og mærkevareudbydere ved analyse af den kolossale mængde af scanner data. Måling af promotioneffekter, prisfølsomhed, reklameeffekter og input til mere effektiv space management er blot nogle eksempler.

Efter en beskrivelse af strukturen i scanner data vil det blive diskuteret, hvilke muligheder der er til stede ved analyse af scanner data alene, og hvilke yderligere muligheder der opstår ved at kombinere scanner data med data fra et husstandspanel.

Scanner data

Hver kasse i supermarkedet er forbundet til en database, som indeholder oplysninger om hver eneste EAN¹-kodet artikel i butikken. Oplysningerne er såvel statiske som dynamiske. Den statiske information vedrører producent, størrelse af artiklen osv. Den dynamiske information refererer til fakta om artiklen, der varierer over tid, som løbende pris og antal solgte enheder siden sidste initialisering af databasen. Det er normalt, at supermarkedskæden sender data fra hver butik til en central computer med faste tidsintervaller, typisk pr. dag. Pr. dag vil databasen i butikken derfor blive initialiseret.

Når ekspedienten fører en vare hen over scanneren, bliver artiklens nummer læst og sendt til databasen, hvor der sker to ting. Først bliver prisen på artiklen fundet i databasen og brugt til beregning af kundens regning. For det andet bliver det faktum, at artiklen er solgt, registreret i databasen ved at lægge 1 til det totale antal solgte enheder for denne artikel. Endvidere indeholder databasen oplysninger om

det totale salg i kroner af den pågældende artikel.

Scanner data vil kunne foreligge på forskellige aggregeringsniveauer. Det mest typiske for øjeblikket er, at opgørelsen sker pr. EAN nummer pr. dag pr. butik. En record i databasen vil i dette tilfælde indeholde oplysningerne :

- EAN koden
- pris
- antal solgte enheder

Man kan naturligvis også forestille sig aggregering på tværs af butikker for en given region eller aggregering af dagsdata til uge- eller månedsdata. Endelig kan man forestille sig pooling på tværs af EAN-numre fx. for at opnå et tal for salget i en given kategori pr. dag.

Aggregering bør finde sted med grundig omtanke og kun hvis de enheder, der summeres, er homogene mht. marketingaktiviteter. Hvis der fx. er en 2-ugers periode pr. måned med promotion, og data summeres fra ugeniveau til månedsniveau, vil al information om promotioneffekt gå tabt. Derimod vil summering til to ugers perioder (promotion / ikke-promotion) være acceptabel, da de summerede perioder marketingmæssigt er homogene. Samme problem kan opstå ved at summere data over butikker med forskellige marketingaktiviteter. Man kan endelig forestille sig bias selv i data på butiksniveau, hvis butikkens ugentlige opgørelser af scannerdata ikke følger ugerne for promotionaktiviteter. Bemærk at homogenitet ikke nødvendigvis kræver homogenitet i respons på marketingaktiviteterne. Der er generel konsensus om, at homogenitet i aktiviteter

er vigtigere end homogenitet i respons (Allenby&Rossi, 1991).

Man kan også forestille sig, at det vil være interessant at opgøre salget på timebasis for at optimere åbningstiderne. Med den udvidede åbningstid om aftenen i Danmark er det selvfølgelig værdifuldt at kunne opgøre, hvad der sælges på disse ydertidspunkter.

Som beskrevet ovenfor er der typisk én record i databasen pr. EAN nummer, men det kunne også være interessant at organisere data, så der var én record i databasen pr. indkøb med en registrering af, hvilke artikler, der er købt på samme tidspunkt. Analyser af denne type data kaldes for kurveanalyser (Julander,1992). Ved at vælge det samlede indkøb som analyseenhed, bliver analysen mere forbruger- eller detaillistorienteret, da indkøbskurve er hvad forbrugere køber, og hvad detaillister sælger. EAN-nummeret som analyseenhed er det mest oplagte for mærkevareudbyderen.

Kurveanalyser kræver ikke, at man har et husstandspanel, men det vil naturligvis være interessant at vide, hvem der kører af sted med indkøbsvognen. Scanner data vil kunne bruges til at etablere en hel indkøbshistorie pr. husstand, hvis man indfører et forbrugerpanel. Når et medlem af panelet handler i supermarkedet føres et kort igennem, der identificerer paneldeltageren sammen med det indkøb vedkommende foretager. I Danmark vil FDB fuldt lovligt kunne udnytte sine medlemskort til dette formål, da der ikke er tale om kreditkort, men det vil FDB ikke, og den officielle begrundelse er, at FDB ikke i aftalen med kunderne har angivet, at man vil udnytte kortet på den måde².

Den scannede datamængde er kolossal³, og blot det at bringe data på en brugbar form er bestemt ikke noget trivielt problem. Udvikling af systemer, som kan transformere disse data til beslutninger, der kan give konkurrencemæssige fordele er en opgave, der både kræver store investeringer såvel i mennesker som i selve systemet.

Der foreligger flere ekspertsystemer til analyse af scanner data fx. et system ved navn INFER (Rangaswamy et al.,1991). Der er både fordele og ulemper ved et ekspertsystem. Når det engang er udviklet, er det nemt at reproducere og distribuere. Et ekspertsystem erstatter den menneskelige ekspert og er mere konsistent forstået på den måde, at ekspertsystemet altid kommer frem til samme resultat med samme informationer til rådighed. På minussiden tæller det, at ekspertsystemet vil være sensitivt over for ændringer i markedsbetingelser. Vedligeholdelse af et ekspertsystem vil være meget kostbart. Da viden om alle marketingforhold stadig er ukomplet, er det svært at udvikle et ekspertsystem, der kan indeholde alle de nødvendige regler til at afspejle markedets dynamik.

To firmaer og to produkter

Detailhandlere i hele verden sælger scanner data på butiksniveau pr. uge til enten A.C. Nielsen eller IRI (Information Resources Inc.)⁴. A.C. Niensens produkt hedder SCANTRACK, mens IRI's produkt hedder INFOSCAN. A.C. Nielsen er ene om markedet i Danmark og dækker ca. 18% af omsætningen i dansk dagligvarehandel⁵. Alle dagligvarekæder er repræsenteret bortset fra ALDI, og butikkerne

er udvalgt efter størrelse, geografi og andre forhold, der skulle sikre repræsentativitet. De scannede data suppleres med besøg af optællere, der registrerer eksponering af varer, displays og andre promotionaktiviteter. Informationen er nedbrudt på EAN nummer samt summeret på varemærke, producent og varegruppe.

IRI er repræsenteret i 15 forskellige europæiske lande fx. Sverige, hvor de har indgået en strategisk alliance med GfK, og hvor de har en andel af markedet for scanner data.

A.C. Nielsen har ikke noget husstandspanel i Danmark. Det tilbyder GfK derimod. Panelet består af 2300 husstande, der jvf. GfK⁶ er udvalgt på en sådan måde, at de i henhold til demografiske kriterier er repræsentative for husstande i Danmark. Deres produkt kaldes for ConsumerScan,

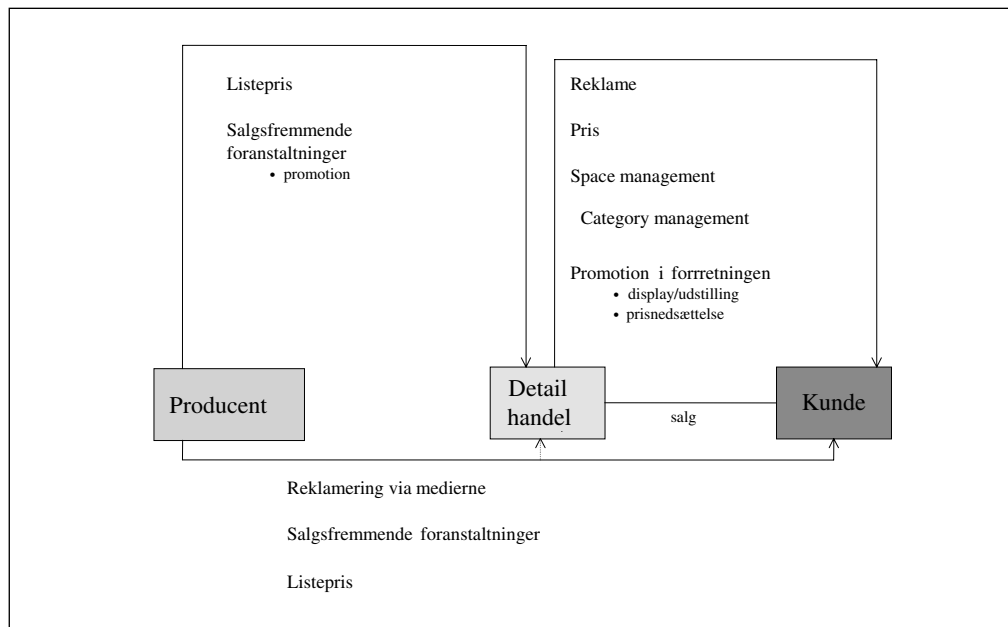
og resultater herfra leveres til de fleste mærkevareleverandører i Danmark.

Muligheder i de rene scanner data

Generelt vil analyse af scanner data kunne anvendes til at vurdere reaktionen på marketingaktiviteter i hele distributionslinien som vist i figur 1. Figuren er tilpasset efter en figur i Little(1991).

Producenten kan følge konsekvenserne af sin egen og konkurrenters promotion, prispolitik og strategier for nye produkter. Analyse af scanner data vil kunne gøre det muligt i tide at reagere både taktisk og strategisk og justere promotion og produktion i henhold til analyseresultaterne. Detaillisten kan fx. anvende data til disponering af indkøb, sortimentstilpasning, ved prisforhandlinger med producenten og ved disponering af hyldeplads.

Figur 1. Distributionskanalen og marketingaktiviteter.



I det følgende vil de mest oplagte anvendelsesområder for de rene scanner data blive kommenteret.

A. Vurdering af promotionaktiviteter

I Davidson&Stacey(1997) afrapporteres resultaterne fra telefoninterviews gennemført med 56 dagligvareproducenter i USA om den relative betydning af 15 marketingforhold og i hvilket omfang scanner data bliver brugt til at vurdere disse forhold. Evaluering af promotionaktiviteter⁷ opnår højeste betydningsscore, og scanner data anvendes intensivt i den sammenhæng. Promotionaktiviteten evalueres typisk ved at gennemføre et såkaldt baselinestudie med det formål at måle, hvad der kan sælges uden promotion (baseline) og hvilket ekstra salg, der kan henføres til promotionaktiviteten. Det er naturligtvis derfor, at A.C. Nielsen supplerer de rene scanner data med information om displays etc. På basis af salg i uger uden promotion beregnes en baseline efter en eksponentiel smoothingsmodel, der har den indbyggede egenskab, at den giver mest vægt til de seneste ugers salg. Ved at sammenligne forskellen mellem aktuelt salg og baseline-salg i en given uge og med registrering af den gennemførte promotion, kan resultaterne anvendes til at vurdere effektiviteten af forskellige promotionaktiviteter. I den forbindelse er det dog værd at huske, at forskellen mellem aktuelt salg og baselinesalg er et estimat og ikke en observation. Det er endvidere vigtigt at understrege, at mersalg ikke nødvendigvis betyder forøget profit.

Problemerne med baselinestudier knytter sig til,

- at loyale kunder køber ekstra ind i uger med promotion
- kunder skifter butik som reaktion på promotion
- kannibalisering af søsterprodukter
- konkurrenters promotion

Disse forhold vil alle have en tendens til, at baselinesalget bliver undervurderet med den anvendte metode, så effekten af promotion bliver overvurderet. Hvis effekten af disse udeladte forhold er den samme uafhængigt af karakteren af promotion, så vil producenten stadig kunne anvende data til at vurdere den relative effektivitet af promotionaktiviteterne og dermed til en allokering af et givet promotionbudget, men ikke til optimering af størrelsen af det totale promotionbudget.

B. Vurdering af prispolitik

Scanner data vil kunne anvendes og bliver anvendt også i Danmark til bestemmelse af priselasticiteter. Jvf. Bucklin&Gupta (1999) er det dog en udbredt opfattelse, at forbrugerne ikke reagerer på prisændringer i et givet interval, men der eksisterer en prisgrænse, som vil føre til en markant reaktion fra forbrugerne. Det vil naturligvis være af interesse, hvis man via analyse af scanner data kunne finde frem til, hvor stort det maksimale gap mellem eget produkt og konkurrerende produkt kan være, uden forbrugerne reagerer.

Priselasticiteter beregnes typisk på basis af data på butiksniveau. Der foreligger en tidsserie af observationer for en række butikker, og dermed gennemføres en såkaldt tidsserie/cross-section regressionsanalyse. Modellen er typisk formuleret som vist nedenfor :

$$\ln(\text{Solgt_mængde}) = \beta_0 + \beta_1 \ln(\text{Pris}) + \beta_2 \text{Prom}_1 + \beta_3 \text{Prom}_2 + \beta_4 \text{Butik} + \beta_5 \text{Uge} + \varepsilon$$

hvor $\text{Prom}_1, \text{Prom}_2$ er dummy-variable, der markerer om bestemte promotionaktiviteter har været anvendt i en given uge i en given butik⁸. β_1 svarer til priselasticiteten for den analyserede vare, der med den valgte modelspecifikation antages konstant på tværs af alle uger og alle butikker. Modellen vil kunne udbygges med led, der tager højde for, at priselasticiteten ikke er konstant. Der kan også inkluderes konkurrerende priser og registreringer af konkurrenters promotion i modellen for de mest betydningsfulde konkurrenter. Der har dog været problemer i visse studier med at opnå elasticiteter med de korrekte fortegn baseret på scanner data på butiksniveau især for krydspriselasticiteter (Blatberg&Wiesniewski(1989)). Hvis man finder statistisk signifikante priselasticiteter med forkerte fortegn er modellen ikke korrekt specificeret. Det kan skyldes, at man har glemt en vigtig variabel. Det kan også skyldes, at funktionsformen er forkert, men det er dog meget usandsynligt, da den log-lineære model kan opfattes som en lineær approksimation til enhver bagvedliggende funktionsform. Boatwright et al.(1999) er det nyeste bidrag til denne diskussion og anbefaler, at estimationen gennemføres med en Bayesiansk estimationsprocedure, der tager hensyn til fortegnsinformation, så man undgår ukorrekte fortegn i stedet for den traditionelle OLS⁹-regression.

Bestemmelse af acceptable prisforskelle mellem eget produkt og konkurrerende produkt foregår i praksis¹⁰ på følgende simple måde :

- For alle butikker af samme type indsamles information om samlet salg i butikken, salg af det analyserede produkt og den normale prisforskel mellem produktet og det konkurrerende produkt
- Samlet butikssalg og salg af produktet aggregeres for alle butikker, der har samme normale prisforskel.
- Det beregnes hvor stor en andel af det samlede salg som salget af produktet udgør¹¹.
- Plot salgsandele som funktion af forskelle i normalpris. Vurder størrelsen af den maksimalt acceptable prisforskel ved at se på kurven.

Metoden kan give et førstehåndsindtryk af størrelsen af den maksimalt acceptable prisforskel, men baserer sig på data på tværs af butikker. Der er mange forskelle mellem de forbrugere, der besøger butikker i fx. Hellerup og Hanstholm, så mere stabile resultater for den maksimalt acceptable prisforskel vil kunne opnås ved analyser af data, der kommer fra samme butik, men hvor prisforskellen er ændret over tid.

For en diskussion af mulighederne i scannerdata til måling af kortsigtede og langsigtede reklameeffekter og til sortimentstilpasning henvises til Bucklin& Gupta(1999).

Kurveanalyser

Resultater fra kurveanalyser kan anvendes til optimering af pris- og promotionstrategi på tværs af en række produktkategorier. Det er interessant både for en producent og en detaillist. Analyse af indkøbskurvens samlede indhold kan fx. blive anvendt til at vurdere, hvilke varer, det vil være mest fordelagtigt at sætte på tilbud, da salget af

den pågældende vare typisk føder et salg af et sæt af andre varer. Der kan imidlertid være mange årsager til at to varer optræder i samme indkøbskurv, hvor komplementaritet er én af disse. Det kan også skyldes, at varerne har samme indkøbsfrekvens, men også en stor mængde af uobserverede faktorer som fx tidspres på forbrugeren, der gør, at hun samler en mængde indkøb. Komplementaritet giver detaillisten en vis kontrol over forbrugernes indkøb, mens de øvrige faktorer er langt mindre kontrollerbare. Det vil endvidere være oplagt ved analyse af kurve at tage hensyn til forbrugerheterogenitet. Der er god grund til at tro, at pris- og promotionaktiviteter vil påvirke forbrugernes indkøb af komplementære produkter forskelligt og dermed

vil forbrugerheterogenitet være en betydningsfuld faktor til forståelse af den sande sammenhæng mellem forskellige produktkategorier. Hensyntagen hertil vil selvfølgelig kræve, at scanner data suppleres med et husstandspanel.

Det er almindeligt at opstille tabeller, der viser hvilke produktkategorier, der typisk findes i samme indkøbskurv. Som eksempel herpå gengives nedenstående tabel, som stammer fra en offentlig tilgængelig præsentation på GfK's hjemmeside og derfor er beregnet på basis af data fra deres husstandspanel, men intet er til hinder for at opstille en lignende tabel på basis af scanner data. Tabellen viser i hvor mange % af de analyserede kurve, man finder varer fra par af kategorier.

Figur 2. En tabel fra en traditionel kurveanalyse.

	Category 2	Category 3	Category 4	Category 5	Category 6	Category 7	Category 8	Category 9	Category 10	Category 11	Category 12	Category 13	Category 14	Category 15	Category 16	Category 17
Category 1	30%	14%	11%	14%	8%	11%	14%	43%	0%	14%	22%	11%	16%	9%	16%	11%
Category 2	100%	9%	12%	16%	14%	13%	13%	13%	11%	13%	18%	13%	14%	14%	14%	12%
Category 3		100%	19%	16%	14%	12%	14%	12%	21%	20%	13%	12%	21%	15%	20%	11%
Category 4			100%	9%	23%	24%	11%	23%	19%	18%	20%	20%	25%	17%	15%	14%
Category 5				100%	13%	10%	9%	22%	11%	10%	18%	18%	9%	12%	12%	12%
Category 6					100%	13%	15%	18%	6%	14%	13%	20%	13%	22%	28%	13%
Category 7						100%	16%	15%	0%	12%	12%	12%	16%	15%	11%	10%
Category 8							100%	14%	18%	14%	15%	14%	15%	16%	14%	16%
Category 9								100%	29%	15%	16%	19%	19%	14%	17%	18%
Category 10									100%	4%	12%	17%	27%	6%	19%	8%
Category 11										100%	13%	29%	21%	17%	16%	15%
Category 12											100%	11%	14%	20%	14%	14%
Category 13												100%	30%	18%	20%	15%
Category 14													100%	21%	14%	15%
Category 15														100%	20%	17%
Category 16															100%	13%
Category 17																100%
Category 18																
Category 19																
Category 20																
Category 21																
Category 22																
Category 23																
Category 24																

Det besnærende ved den skitserede analyseform i figur 2 er, at den er meget simpel, men det problematiske er, at man ikke kan skelne de forskellige årsager til, at det fælles køb er foretaget. I Manchanda et al(1999) præsenteres en model, der tillader afhængighed mellem de valgte kategorier og opdeler i effekter fra kontrollerbare og ukontrollerbare drivere for køb. Modellen kræver gentagne observationer fra samme husstand, så det er altså nødvendigt med et panel for at benytte modellen.

Scanner data kombineret med data fra et husstandspanel

Analyse af scanner data på husstandsniveau giver enestående muligheder for at forstå forbrugerens indkøbsadfærd og udlede implikationer heraf for marketingaktiviteter. Som eksempler kan nævnes :

- Måling af prøve- og genkøbssandsynligheder for nye produkter
- Måling af mærkevareloyalitet
- Måling af graden af forbrugerheterogenitet og udnyttelse af kendskab til segmenter fx. til direct mail kampagner
- Måling af effekten af promotion på forbrugerens tendens til at købe ekstra ind i sådanne uger
- Måling af forbrugerens tendens til mærkevareskift over tid og dermed bidrage til forståelse af markedsstruktur og konkurrencestruktur.

Det store spørgsmål i forbindelse med husstandspaneler er, om de er repræsentative. Der kan være flere årsager til, at et panel ikke giver repræsentative resultater. Der kan for det første være forskel mellem de husstande, der indvilliger og de hus-

stande, der nægter at deltage i et panel. En af årsagerne til en sådan forskel kunne være det store arbejde, der i sin tid var forbundet med det at føre dagbog. I vore dages scannertider, hvor man blot skal føre plastickortet igennem i butikken¹², vil besværet være minimalt for forbrugeren. Det må også forventes at give mere præcise resultater end i dagbogstiderne, og forbrugeren er mere ubevidst om at deltage i et panel. Hun kan selvfølgelig bruge kortet ufuldstændigt, hvilket også vil føre til en bias i resultaterne fra analyser af scannerdata på husstandsniveau. Der er gennemført en række analyser af en evt. bias i paneldata ved at sammenligne priselasticiteter estimeret på butiksniveau med priselasticiteter estimeret fra scannerdata på husstandsniveau (Gupta et al(1996). Konklusionen er, at der er meget lille forskel mellem disse elasticiteter, og husstandsdata derfor vil kunne give detaillister og producer værdifuld information om, hvad de dybereliggende årsager er til, at salget udvikler sig, som det gør.

Modellering af forbrugerens valg af mærkevarer (brand choice) er det mest omtalte aspekt af indkøbsadfærd analyseret med scannerdata fra panel. Det skyldes bl.a. den store betydning som forståelsen af årsagerne til mærkevareskift har for detaillisters og mærkevarerudbyderens profit(Gupta(1988)). Den multinomiale logit-model gengivet nedenfor er blevet standard for analyse af valg af mærkevarer :

$$P_{ist} = \frac{\exp(X_{ist}'\beta)}{\sum_j \exp(X_{jst}'\beta)}$$

hvor P_{ist} er sandsynligheden for at mærke i vælges i butik s i uge t , når marketingvariable som pris og promotion antager værdi-

er som angivet i X_{ist} , og β er vektoren af effekter fra marketingvariablene på valget af mærkevarer. Som modellen er angivet ovenfor antages den at gælde for alle husstande, men det er også muligt at antage forskellige effekter af marketingvariablene for segmenter af husstande. Et problem er, at scanner data fra panelet pr. definition kun vil indeholde oplysninger om priser for den solgte vare og derfor skal suppleres med oplysninger om promotion for den valgte vare og priser for de konkurrerende varer, som blev fravalgt. Erdem et al(1999) har beskrevet en mulig metode til at korrigere estimerne på marketingeffekter for manglende prisoplysninger på konkurrerende varer.

Multinomialmodellen bruges nærmest på rutinebasis i USA (Little(1991)) til at vurdere effektiviteten af forskellige marketingaktiviteter. Proceduren er som følger. Modellen kalibreres først over en periode på fx. 1 år uden en bestemt promoti- onaktivitet. Dernæst introduceres pro- motionaktiviteten og aktuelt salg med promotion sammenlignes med det salg modellen forudsiger. Differencen mellem aktuelt og forventet salg er så et udtryk for promotioneffekten.

Konklusion

Med de kolossale mængder af information, som scanner data indeholder, er der mere end nogensinde brug for værktøjer, der kan trække de generelle mønstre ud af data. I dette papir er det diskuteret, hvorledes data kan organiseres og analyseres, når analytikeren selv definerer, hvilke spørgsmål han ønsker besvaret. De behandlede spørgsmål er de højest prioriterede af mærkevarerudbydere og detaillister. De benyttede

redskaber i industrien i dag er diskuteret. Det er klart at metoderne skal være relativt simple for at finde bred anvendelse. Det er de også, men det er naturligvis også vigtigt at kende begrænsningerne på værktøjernes anvendelse, så det har været et af formålene med denne artikel at påpege disse.

De fulde muligheder i scannerdata vil først blive udnyttet i det øjeblik man kan knytte forbindelsen mellem det registrerede indkøb, og karakteristika ved den husstand, der har foretaget indkøbet, da man kun i dette tilfælde, vil kunne udnytte resultaterne til en mere målrettet markedsføring mod bestemte segmenter af husstande.

Scanner data vil også være et oplagt emne for data mining (Mason(1995)). Data mining vil kunne komplementere den mere traditionelle hypotesebaserede analyse, som er behandlet i denne artikel.

Summary

The article discusses some of the opportunities available to the manufacturer of brands as well as the retailer, to improve the efficiency of marketing activities, using scanner data information.

Noter

¹ EAN (European Article Numbering). EAN koden har 13 cifre, der entydigt identificerer producent og produkt.

² Børsen d. 18/5 2000

³ jvf. EAN Danmark har Arla Foods til eksempel rådighed over 800000 EAN numre.

⁴ Data fra A.C. Nielsen i Danmark offentliggøres af kontraktlige årsager med 4 ugers intervaller. Man kan ikke se resultater for den enkelte butik, men fx. for BILKA, OBS, FØTEX, KVICKLY og for hele discountkæden under ét.

⁵ Personlig kommunikation med Malene Olsen, A.C. Nielsen – AIM.

⁶ <http://www.gfk.dk>

⁷ ved promotion forstås ethvert finansielt incitament fra producent til detaillist til at nedsætte prisen på en artikel, placere produktet på en fremtrædende plads i butikken eller markedsføre artiklen i en tilbudsavis, avisreklame eller lignende.

⁸ Der kan selvfølgelig være mere end to af disse dummy-variable, hvis der er mere end to forskellige promotionaktiviteter, man ønsker at tage højde for.

⁹ OLS står for Ordinary Least Squares – mindste kvadraters metode og er standarden i ethvert regressionsprogram fx. i SPSS eller SAS.

¹⁰ I hvert fald i USA jvf. Bucklin&Gupta(1999)

¹¹ Dermed korrigeres for forskellig butiksstørrelse

¹² fx. A.C. Nielsens ERIM panel og IRI's BehaviorScan. (De findes ikke i Danmark)

Litteratur

Allenby, G.M. & Rossi, P.E.: There is no aggregation bias: Why macro logit models work. *Journal of Business and Economic Statistics*, 9, 1-11, 1991.

Blattberg, R.C. & Wiesniewski, K.J.: Price-induced patterns of competition. *Marketing Science*, 8, Fall, 291-309, 1989.

Boatwright, P.; McCulloch, R. & Rossi, P.: Account-level modeling for trade promotion: An application of a constrained parameter hierarchical model. *Journal of the American Statistical Association*, 94, 448, 1063-1073, 1999.

Bucklin, R. & Gupta, S.: Commercial Use of UPC scanner data: industry and academic perspectives. *Marketing Science*, 18, 3, 247-273, 1999.

Davidson, T. & Stacey, A.E.C.: 1996 marketing analysis and modeling survey. Summary report, Prevision Corporation, Wellesley, MA, 1997.

Erdem, T. Keane, M.P. & Sun, B.: Missing price and coupon availability data in scanner panels: Correcting for the self-selection bias in choice model parameters. *Journal of Econometrics*, 89, 177-196, 1999.

Gupta, S. Chintagunta, P.; Kaul, A. & Wittink, D.R.: Do household scanner data provide representative inferences from brand choices: a comparison with store data. *Journal of Marketing Research*, 33, November, 383-398, 1996.

Gupta, S.: Impact of Sales Promotions on When, What and How much to buy. *Journal of Marketing Research*, 25, November, 342-355, 1988.

Hawkes, W. & Smith, R.: Improving Consumer Price Measurements through the use of scanner data and market segmentation. <http://www.cf.ac.uk/carbs/conferences/mi99/hawkes.html>, 1999.

Julander, C.R.: Basket analysis: a new way of analyzing scanner data. *International Journal of Retail and Distribution Management*, 20(7), 10-18, 1992.

Little, J.D.C.: Operations research in industry: New opportunities in a changing world. *Operations Research*, 39, 4, 531-542, 1991.

Manchanda, P.; Ansari, A. & Gupta, S.: The "shopping basket": A model for multicategory purchase incidence decisions. *Marketing Science*, 18, 2, 95-114, 1999.

Mason, B.: Data mining: Exploring the unknown. *Discount Merchandiser*, 35, 10 T58-T59, 1995.

Rangaswamy, A.; Harlam, B. A. & Lodish, L.M.: INFER: an expert system for automatic analysis of scanner data. *International Journal of Research in Marketing*, 8, 29-40, 1991.