MODEL-ASSISTED ANALYSIS, SIMULATION AND FORECASTING WITH CONSUMER PANEL DATA

*by Raimund Wildner and Birgit Scherübl**

ABSTRACT

This paper starts out by discussing the marketing models which have been available up to present, then goes on to describe a new model which is based on scanner data from a consumer panel. The data input and the model structure are described, the quality of the model is discussed and some possible applications are indicated. Finally the paper describes the limitations to the model as well as possible further developments.

* ***Dr. Raimund Wildner,*** *e-mail: raimund.wildner@gfk.de*
Birgit Scherübl, *e-mail: birgit.scheruebl@gfk.de, both at GFK-Nürnberg e.V., Nordwestring 101, D-90319 Nürnberg*
Originally published in GfK Jahrbuch der Absatz- und Verbrauchsforschung Vol. 51 (2005), No. 4, pp. 332–354.

1. *Problem definition*

Marketing and distribution are continually confronted with the question of what the impact of a change in the marketing mix will be on market share. How have price and market share moved? The answer to this question is certainly interesting. But more important is the answer to the question of what the market share is to be expected if prices are reduced by 5 %, or better still, at which price the brand-contribution will be maximized.

The answers to such questions are expected from **marketing mix models**. The same **principle** underlies all such models. Initially a correlation is calculated between input variables (e. g. price, distribution, advertising expenses) and an output variable (e. g. market share). Once this correlation has been established and checked for quality in terms of content and statistical validity, **three** main **types of question** can be answered, which can each best be explained by using an example:

1. **Analysis:** What caused an increase in market share? Very often during the course of two time-periods (for instance first six months of 2005 compared with the first six months of 2004) there have been simultaneous changes in a very large number of input variables. Marketing mix models make it possible to allocate the whole change to the various input variables.
2. **Simulation:** How would the market share change, if the price were raised by 5%? Changes in the values in marketing mix inputs can be entered into the model. Using the model's equations, an estimate of the effect on the target value is made. In doing this it is possible just to change one variable (price in this example) or alternatively to alter all the input variables simultaneously, so that complete scenarios can be checked for their impact.
3. **Forecast:** What market share can be expected with the planned marketing mix? If a future scenario is entered, then the simulation becomes the forecast. For competitive activities, assumptions have to be made (for instance that the competition operates in an optimum manner).

Marketing mix models were developed **decades ago** (*Amstutz* 1967; *Lavington* 1972; *Klenger* and *Krautter* 1973; *Little* 1974), but were not put to practical use for a long time since the models were not compatible with the available data. At the beginning of the 1990s a model was developed, the GfK BrandSimulator based on retail audit data, which was compatible with traditional retail panel data (*Wildner* 1990 and 1991) and which was also used in practice (*Vossebein* and *Wildner* 1992). Later it was converted for the use of retail scanner panel data (*Wildner* 1994). Retail scanner panel data are also used by other market research companies, i. e. IRI (www.infores.com) and A.C. Nielsen (www.acnielsen.com) for marketing mix modelling.

The current paper discusses a new type of model which is based on consumer scanner panel data. In the second chapter, the characteristics of modelling using consumer panel data are compared to those using retail panel data. In the third chapter the model is explained in more detail, applications and two case studies are described in the fourth chapter. Finally in the last part, the limitations and possible future developments are discussed.

2. *Retail and consumer panel data as input to marketing mix models*

2.1 *Retail panel data as input*

Retail scanner panel data are collected from the scanner tills in selected retail outlets. For each article, the volume sales per week and the unit price are recorded automatically. Additionally in a sub-sample, details of retail promotions are collected (*Günther et al.* 1998 pp. 69 f). There is therefore one data point per shop and per week. Prices are entered on an unweighted basis. Retail promotions are coded with so-called „dummy-variables“, where data relating to shops and weeks with the relevant promotions are coded with 1, while all other data is coded with 0.

That means that all those variables which affect individual retail outlets are represented in a very differentiated way. On the other hand, factors which have a **direct impact on the consumer** but do not target the shops individually (such as advertising) can only be taken into account **in very general terms**, with all data records from shops in the relevant area and in the relevant weeks being given the same code (for instance on advertising pressure).

In reality it is to be assumed that the advertising effect is not the same in all shops. Some shops for instance are mainly used by people in employment, while others tend to be more used by elder people in their retirement. The pattern of TV viewing by the two groups is very different and they are therefore exposed to different advertising.

But if, as a result of the lack of better information on advertising, all shops are handled in the same way, then the relevant variable will contain error. It is known from econometrics that random error in input variables leads to underestimates of their impact (on the „error in exogenous variables“ model, see *Schneeweiß* 1974, p. 223 as an example).

Additionally, all retail scanner panel models actually model all shops independently of each other. They take into account if a promotion takes place in a certain shop. But no information is available about promotions in other nearby shops. Obviously promotions lead to consumers changing shops (that is why after all retailers do them). That means that the increase in the number of consumers resulting from migration between shops is modelled but not the decrease in other outlets. This results in the effect of retail promotions tending to be over-estimated.

It can therefore be concluded that models which are based on retail scanner panel data tend to **overestimate** the **effects of promotions** and **of price changes**, while they tend to underestimate the effect of marketing activities, such as advertising or the distribution of samples where there is a **direct contact with the consumer**.

2.2 *Consumer panel data as input*

The situation is different with those models which are based on consumer panel data. The data for the GfK Household Panel is gathered with a hand scanner, called the „electronic diary" (*Günther et. al.* 1998, pp. 38 f) or since 2005 with a barcode scanner and a PC. The following data is collected:

- precise article details (by reading the EAN-Code (1) with the CCD-scanner in the electronic diary device or with the hand scanner or by describing it using a dialog).
- the number of units of each article bought by entering the data, using the keyboard (of the electronic diary device or the PC).
- Price paid (entered via the keyboard).
- information on whether the article was bought as part of a promotion (keyboard).
- the date of purchase and the shop where each purchase was made is entered once via the keyboard for each purchase act, but recorded for each article.

Since data on households is available (a total of 13,000 households up to 2004, 17,000 households in 2005 and 20,000 households from 2006 onwards), all the input variables which affect the individual household, such as advertising, distribution of leaflets and samples can be taken into account in a very differentiated way per household. Furthermore, it is known whether a product which was bought by the household was a promotional item. On competitive products we only know from the purchases by other households about availability, price and whether they were on promotion or not. This information will contain some errors with those retail organizations which are not centrally managed. So concerning the data spectrum, consumer panel data are superior to retail panel data in some respects and inferior in others.

Concerning the modelling of the migration between retail outlets consumer panel data offer a clear advantage. Since we know from a household about the shops he is visiting, this can be done quite rapidly.

Short models, which are based on the Household Scanner Panel are very **well suited** for modelling the effect of those marketing mix inputs, which have a **direct impact on the household** (advertising, leaflets, samples). On the other hand, models based on retail scanner data are very well suited for modelling those effects that take place in the shop, i.e. distribution, prices and promotions.

Kopperschmidt (2005) also emphasises the necessity of integrating TV advertising on the basis of data from individual households, because with „macrodata there is no interface between individual purchase acts and households' exposure to advertising. Therefore there are only limited possibilities for integrating into models theories of advertising based on aggregated data" (referred to on p. 6).

3. *Building a new marketing mix model on the basis of consumer panel data*

3.1 *Objective*

With the new model, named the "**BrandSimulator**" based on consumer panel data, it is intended to achieve the following **objectives**:

- Simultaneous estimation of all brands and all retail outlets. This also includes modelling the migration between brands and retail outlets. It comprises therefore a marketmodel, and not a model for the individual brand. This is important, because it is the only way that the question can be answered whether if a company puts up the price on its product A, its other product B benefits or competitive product C.
- Estimates based on individual purchase acts. This is the only way to construct and to investigate all types of target group.
- Simultaneous modelling of the effect of the most important marketing mix inputs, i.e. price, distribution, promotion and TV advertising. This is the only way that the marketing mix as a whole can be optimised.

In the **literature,** there are numerous models which operate on the basis of household panel data. Admittedly, with them only certain dimensions of a purchase act can be examined. Thus models usually only cover either brand choice, time of purchase or volume (for instance *Gupta* 1988; *Ailawadi* and *Neslin* 1998). *Ailawadi* and *Neslin* (1998) in their model concentrate on the interface between promotional activity and consumption of the category. There is no attention given to the selection of the retail outlet or TV advertising. *Ailawadi*, *Gedenk*, *Lutzky* and *Neslin* (2005) model the interconnection between promotion and stocks at a known time of purchase in a known shop. Using household panel data, *Kopperschmidt* (2005) has developed a model to cover the choice of time of purchase. In this model, he integrates TV advertising at household level. *Fader* and *Hardie* (1996) do not model brand choice but rather explain it from a set of characteristics. The present authors are not aware of any model, based on consumer panel data which integrates brand choice, selection of retail outlet, time of purchase and sales volume into a single model which also takes into account distribution, TV advertising, prices and promotion for the client's own product and also competitive products.

3.2 *Data input*

The BrandSimulator utilises Household Scanner Panel data on a category and for a period of between one and two years (base period) at the level of individual purchase acts. **Certain households are excluded** from the initial data:

1. Households are only used which form part of what is called the "constant data pool", i.e. which have reported without interruption from the beginning to the end of the base period. Only incomplete data is available from the other households. They are therefore not suitable for being used in model-building.
2. Households are only used which have made a minimum of three purchase acts within the base period. Since every purchase act represents a data point, there is too little information from other households for forming solid estimates.

The proportion of a panel which is accounted for by the constant data pool is a constant number (in a well-run panel, it usually lies between 70 % and 80 % per year); the second rule results in the proportion of purchases which can be utilised in the model depending on the category. The more frequently a category is bought, the higher the proportion. Equally with very infrequently bought products, the value can be so low that building a model is not worthwhile. Taking fabric softeners as an example of an infrequently bought category (on average five purchases per buyer per year), the situation in 2002 was as follows:

1. Out of the 13,000 households on the GfK Household Panel at that time, 7,424 bought fabric softeners at least once.
2. Out of these 7,424 households, 5,524 formed part of the constant data pool, i.e. reported continuously through 2002.
3. Out of these 5,524 households, 3,306 households bought at least three times. The purchase acts by these households form the data base.

It can be seen that only 45 % of all buyers of fabric softeners are included in the model. Admittedly these people make 32,281 purchases and account for 72 % of purchase acts by all households and 87% of purchases in the constant data pool. Since household panel clients are accustomed to special analyses being undertaken on the constant data pool, the last figure is the one that matters. The category is therefore well-suited to being used for model-building.

It is necessary in a further step to determine how the **structure of retailing** for the category is to be represented in the model. The smallest unit which is used is called a

Table 1

Key Account list for the category fabric softener and the year 2002		
E-Aktiv (SM)	toom (HM)	Lidl (Disc)
Spar-SEH (SM)	Wal*Mart (HM)	Norma (Disc)
HL (SM)	Kaufland (HM)	Schlecker (DM)
Tengelmann/Kaisers (SM)	Globus St. Wendel (HM)	dm (DM)
E-Neukauf (HM)	real,- (HM)	Rossmann (DM)
Extra (HM)	Penny (Disc)	other drug markets
miniMAL (HM)	Netto (Disc)	Edeka other
Spar-Regie (HM)	Edeka-Discount	Rewe other
E-Center (HM)	Plus (Disc)	Other grocery shops
Marktkauf (HM)	Aldi (Disc)	

SM = Supermarkets, HM = Hypermarkets, Disc = Discounter, DM = Drug markets

"key-account". The various key-accounts should be as different as possible, but internally as homogenous as possible so that as much of the variance as possible can be represented in the model. Additionally they should be relevant to the manufacturer's distribution and neither too big nor too small, otherwise too many differences will be generated, or alternatively the sample variances will be too great. *Table 1* for instance shows the allocation of key-accounts made for 2002 for the fabric softener category.

Additionally, the **structure of the brands or products** has to be determined. In markets where one pack size predominates (for instance ground coffee in 500 gram packs or 100 gram tablet chocolate), the other pack sizes are usually excluded from the analysis. With fabric softeners, 0.75 litre packs actually predominate but not to such an extent that other pack sizes can be ignored. For this reason, all pack sizes are included but the prices are re-calculated on the basis of the standard 0.75 litre pack.

The five largest manufacturers which have market shares between 9 % and 27 % cover something over two-thirds of the market (volume). Two smaller brands with a market share of around 1 % were also included. In addition retail brands and Aldi, which together hold something over 30 % of the market, have an important role, one that has been growing during the period covered in the research. Other brands are not taken into account, because the data base is too small for each individual brand and the formation of an artificial „others“ brand does not seem appropriate. Such a „brand“ would mix high and low price products and differences would be disguised. In any case, the other brands only have together a share of 7.5 % so the error that has to be allowed for is not significant.

Finally it is necessary to determine **which facts are to be taken from which source. Purchase acts** and **prices paid** come from the Household Panel. The **prices of competitive products** in key-accounts are also taken from the Household Panel, although purchases by other households are used. **Distribution** data comes – where available – from the Retail Panel. If no retail panel data is available (for instance on non-audited categorys or from retail channels such as Aldi or Wal*Mart, which do not co-operate with retail audit companies), then distribution data is inferred from the Consumer Panel, with it always being assumed that a product is distributed to a key-account if a purchase act from it has been recorded.

The handling of **retail promotions** was not so simple. Certainly, households provide information on whether an article was bought in a promotion. Checks have, however, shown that the data recording is incomplete, either due to laziness or to memory lapses of the reporting households. Such data has been shown to be unsuitable for modelling. Reconstructing promotions from the data has been demonstrated to be a successful technique (2). Short-term price reductions combined with significant increases in sales are clear indications of retail promotions, allowing the relevant key-accounts in the relevant weeks to be appropriately identified.

As TV-viewership data is not collected in the GfK Household Panel (3), in order to include **TV advertising**, the number of contacts to TV advertisements must be estimated in advance; this done through a **data fusion**.

As already described GfK has detailed information about the buying behaviour of the households of the consumer panel. Furthermore, GfK has detailed information about the TV viewing behaviour from the appr. 5,500 households of the AGF/GfK TV panel which belongs to the AGF (a working group of German TV broadcasters) and is run by GfK. But these are two separate data sets. What is needed is one data set that contains the TV viewing behaviour and the purchase behaviour from the same households. One possibility would be to collect from the consumer panellists both purchase and TV viewing behaviour.

To evaluate the effects of such a **single source approach** in 1995 a test was performed. Four identical structured groups of 500 panellists each were built: The first group only reported TV viewing behaviour, the second group only purchase behaviour using the microcomputer, the third group both TV viewing and purchase behaviour likewise, the fourth group just purchase behaviour, but with a paper questionnaire.

The results were quite clear: The willingness to report both TV viewing and buying behaviour was about half of the willingness to report just TV viewing behaviour. It is obvious that this has negative effects on the representativity of the panels that are hard to control. In addition, the rate of panellists terminating their co-operation increased by about 50 %. Since many panel analyses only can be done for those that show constant cooperation during the period analysed this is equivalent to a drop in sample size. Since panel dropouts show a different structure than the total sample (e.g. they are younger) the sample gets more biased. In addition, both effects increase the costs of recruiting and running the panel (*Anonymous* 1995).

The other possibility to get a data set with both the purchase behaviour and the TV viewing behaviour is a data fusion between GfK's consumer panel and the AGF/GfK Television Panel. This can be described as follows (*Wildner* 2000):

Target of this data fusion is the evaluation of advertising effectiveness. From the total set of TV-viewing data just the information is needed how often and when a household keeping person has seen advertising for the brand and what was the total advertising viewing time. So the TV panel households are donors which give the information needed to the consumer panel households which are denoted as recipients. The data of the household keeping person in TV panel is used because in consumer panel research in practice the purchase behaviour of the household keeping person is reported.

Concerning TV viewing behaviour we have different quality of data. Whereas the TV panel has electronically measured TV viewing data the household panel just have interview data, where people answer questions how often they watch different types of TV programmes (such as criminal story, soap opera etc.) and how often they use TV at certain days and day parts. This is answered on a four point scale "regularly", "often", "seldom", and "never".

First, these data have to be made comparable. For this a procedure is used, first suggested for fusion problems by Roberts (1994). This takes into account that some programmes show significant overreporting, others are underreported. In addition, the subjective viewing time depends on the time available at all. A pensioner might think that he is watching little if he uses TV for four hours per day because he does also a lot of other leisure time activities whereas an employed mother spending two hours a day watching TV will think that is a lot because it takes nearly all her leisure time.

So in the first step groups are formed according to the working and family status. Then viewing time in TV panel is ranked in descending order. Now if 20 % of the persons of a group of consumer panel declare they are watching regularly at a certain day time, those 20 % of the TV-panel in the same group, that watch at this time most, get the same code. Likewise is done with "often", "seldom", and "never".

Next a regression is computed in the TV panel, where number of commercials of the campaign seen is the dependent variable and the viewing variables together with sociodemographics are the independent variables. Normally we get an R^2 of 60 % to 70 %.

The resulting regression equation is then applied to the households of the consumer panel. By this we get per consumer panel household an estimation of the number of TV spots the household keeping person has seen. This estimation on the consumer panel side and the real figures in TV panel are used to match the panel members. Then the advertising contacts from TV panel are transferred to the consumer panel data base.

3.3 *Model structure*

3.3.1 *Basic structure*

The modelling is based on the individual purchase acts as dependent variable. So that the importance of each of the marketing mix inputs can be assessed in an appropriate way, each purchase act is disaggregated into **four decisions**:

- When is the purchase made? – the question refers to the point in time when the purchase was made.
- Where is the purchase made? – the question refers to the key-account where the purchase was made.
- What is bought? – the question on the brand or the product which was bought.
- How much is bought? – the question on the amount bought.

For each decision there is one sub-model. The four sub-models are finally combined again to purchase acts and purchase volumes. For the succeeding simulations and forecats these are converted into **market shares** and **sales volume**. Besides the marketing mix values there are additional influences being modelled for each sub-model. For every sub-model the relevant variables of influence are listed below. The summary of individual purchase acts results in the rather interesting aggregated values, like purchase acts for one brand in one week and in one key-account. However, if finally only summarised purchase acts are interesting, it is anyway important to model on the basis of individual purchase acts. This is the only way to create any summarised result depending on the question.

3.3.2 *Modelling of the point in time of the purchase act*

When the purchase is made depends on:

- the category price, with only those products which are relevant to the consumer being taken into account

- the promotion share in the category, insofar as it is relevant for the consumer
- seasonality (calculations of weekly sales seasonality, based on at least three years of data). Calendar issues, such as for instance the changing date for Easter have to be allowed for
- time since last purchase: the longer the time since the last purchase within the category, the more likely a purchase becomes
- volume bought at last purchase: the less bought at the last purchase, the greater the likelihood of a further purchase.

3.3.3 *Modelling of the retail outlet*

If the model shows from the time of purchase, that in principle during a given week a purchase was made, then in a second step it has to be determined **where, i. e. in which key-account** the purchase was made. This is influenced by:

- key account (see below) utilities
- benefits from the size of the assortment. This is reflected by the number of products which are in the consumer's relevant set and are in distribution. A product is in the relevant set if it was bought at least once in the base period
- the price level, where the prices for those products are taken into consideration which are relevant to this household and in distribution in the key-account concerned, and are also in the consumer's relevant set in other key accounts. As with a product, a key-account is in the relevant set if at least one purchase was made there in the base period
- retail promotions on the products within the relevant set in the key-account concerned and among relevant competing key-accounts.

Key-account utilities also require explanation:

It is certainly plausible that just on the basis of differences in travel-time, every potential key-account will be preferred to different extents by different households. Additionally, preferences can change as a result of positive or negative experiences – preferences are therefore dynamic.

For each household at least three purchase acts are now available (otherwise the household would be excluded, see paragraph 3.2). It is not possible to estimate from three purchase acts an individual preference which changes through time, because the number of parameters which have to be estimated significantly exceeds the number of data points. For this reason, a different route is followed:

For this purpose, it is assumed that each household prior to a purchase act within a category is at a certain **level in the key-account relationship** structure. The levels in the key-account relationship structure comprise:

- non-buyer
- test-buyer
- repeat buyer
- loyal buyer.

With each purchase in the key-account concerned, the household moves up one level, while with each purchase in a competing key-account, it moves down one level, as long as another level is available. A household which has not made a purchase in a key-account for some time, is accordingly classified as a non-buyer, but if it has repeatedly made purchases and has not bought anywhere else, then it is a loyal buyer. This is how the dynamics of preferences are handled.

A parameter is then estimated for each key-account and loyalty level, which reflects the preference applying to it and which can be interpreted as a utility value. It is also assumed that a key-account has the same level of preference among all those people who are on the same relationship level with that key-account.

The **multinomial logit model** was chosen as the form of sub-model (as also for the other sub-models), because this type of model is particularly well-suited to setting out purchase decisions (*Train* 2003, pp. 41 ff). This sub-model can be represented as follows:

$$p_{a,AKS_{h,w}} = \frac{1}{1 + \sum_{d \neq a} e^{-\left[\sum_{m=1}^{M} \alpha_m \cdot \left(f_{m,a,w,h} - f_{m,d,w,h}\right)\right]}}$$

where:

$p_{a,AKS_{h,w}}$:	probability that key-account a will be chosen in the relevant account-competitive situation $AKS_{h,w}$ by household h in week w. In doing this, an account-competitive situation is defined by the marketing mix situation of the products which are relevant for the household in the relevant key-accounts.
d:	Index for key-account. The summand is present in all relevant key-accounts except for the one currently being examined.
m:	index for marketing mix factors (m = 1,2,...,M)
f:	factor of influence. The form it takes generally is dependent on account a or d, on week w and in household h.
α_m:	parameter which is to be estimated, and which defines the strength of influence exercised by f_m.

3.3.4 *Modelling of the bought brand*

Once the fact of a purchase has been determined and it has been further modelled in which key-account the purchase takes place, it has to be determined **what** is bought, i.e. **which out of the products from the household's relevant set that are in distribution** are selected. This is modelled depending on the following factors, where both the product and the household's relevant competitive products are taken into account:

- distribution
- price
- retail promotions
- brand utilities
- the household's propensity to change brands
- the number and timing of contacts with to TV advertisements prior to the purchase act.

The generation of **prices**, of **distribution** data and data on **promotions** has already been discussed in paragraph 3.2. By analogy with the key-account **utilities**, the brand utilities are modelled at four levels of customer loyalty. A household's propensity to change brands is estimated from the number of different brands bought in the base period as well as a global loyalty parameter, which depends on the category.

For the **modelling of TV advertising** – as already described in paragraph 3.2 – a data fusion with the TV Audience Panel is necessary. When this has been done, then estimated values are available for how many TV advertisements and when the household could see for the brand being considered and the brands competing with it. This enables the effective number of advertising exposures at the time of the purchase to be modelled. For this purpose, the following parameters are estimated:

- a memory factor, which weights the advertisements downwards, the longer ago they ran.
- a minimum value, above which the advertising has an effect. It is therefore possible that the advertisements will initially have no impact, because the message has still to be learned.
- a maximum value, above which further advertising will have no effect, because the message has been learned and is still relevant.

It is also assumed that the utility value of a product correlates linearly with the number of effective advertising contacts, although the linearity factor depends on the level of customer loyalty. The linearity factors as well as the three parameters used for assessing the effectiveness of advertising are estimated in such a way that the influence of the advertising can be maximized.

3.4 *Assessment of the quality of the model*

Criteria for assessing the **quality of the model** comprise:

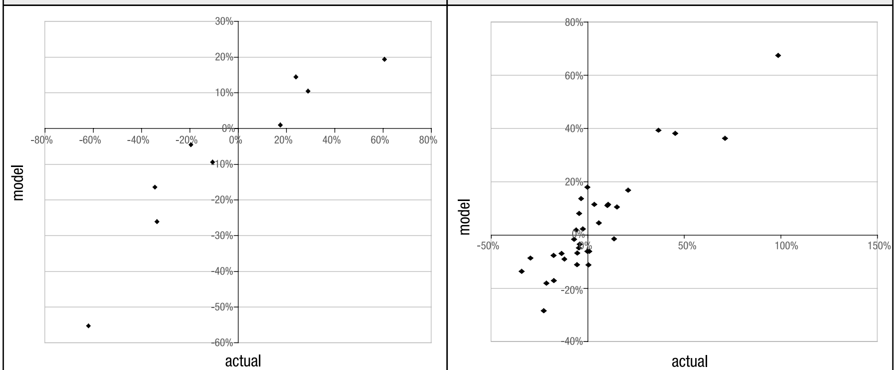
- **model fit**, which provides information on how well the purchasing behaviour as modelled reflects actual purchasing behaviour
- **validation**: this involves estimating the model parameters using data from a base period (in the example of fabric softeners in 2002) and then the actual marketing mix for the time of the forecast (for instance 1-9/2003) is entered. The quality is a result of the fit between the **real** and **forecast** trends in market share.

It is very easy to achieve a good model fit. It generally goes up if additional variables are included in the model, even if these variables in fact cannot provide explanations (so-called „overfitting“). But a reasonable validation result is much more difficult to achieve. The quality of the validation normally declines if additional variables which have little power to explain are added in. Validation is accordingly the much tougher criterion and is to be preferred to model-fitting.

Figure 1

Validation for Fabric Softener and Ground Coffee Categories

fabric softeners	ground coffee



fabric softeners	ground coffee
R^2 = 86.5%	R^2 = 82.7%
base time period: 2002 forecast time period: 1-9/2003	base time period: 1999 forecast time period: 2000

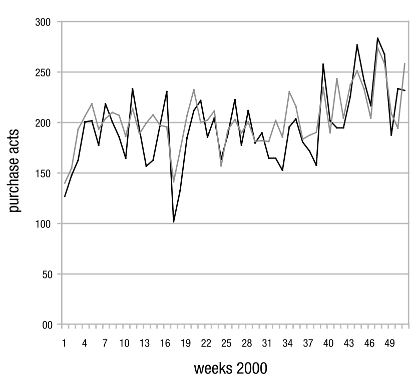
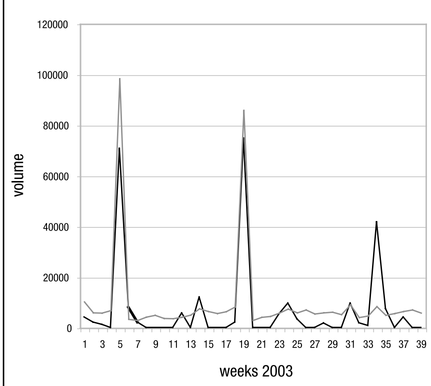
The diagrams show the forecast changes in market share compared with the real ones for the brands included in the model.
Source: own calculations

The validation can be done for the forecast period as a whole. *Figure 1* illustrates this type of validation for nine brands in the fabric softener market (base time period 2002, forecast period 1-9/2003) and for 32 brands in the grounded coffee market (base time period 1999, forecast period 2000). It shows that the model is able to explain over 80 % of the change.

Validation can also be undertaken by looking at individual brands or key-accounts and comparing the market share trends as modelled with what actually happened. *Figure 2* illustrates a validation of this type using a chocolate bar brand in the whole market, together with another validation which shows a brand of fabric softener in one key-account. It shows that the model generates a good result in terms of validation, even if there has been a decline at key-account level.

Figure 2

Validation for a Chocolate Bar Brand and for a Fabric Softener Brand in a Key Account

Chocolate Bar Brand X Total	Fabric Softener Brand Y in Key Account Z
	
real / model	real / model
base time period: 1999 forecast time period: 2000	base time period: 2002 forecast time period: 1-9/2003

The diagrams show the forecast purchase and volume changes compared with real ones.
Source: own calculations

4. *Applications for the new model*

So far the model is run by the French company Marketing Scan for their BehaviorScan test market and by German company GfK and the Dutch company GfK each for the consumer panel.

In fact, the number of **potential applications** is almost limitless. It is possible to examine the effects of as many marketing scenarios as could be desired among a large number of customer groups. Even a very large set of Excel-tables or Powerpoint graphics would only be able to cover them very incompletely. A better solution to the problem will be provided if a **program** can be generated with which the client can himself undertake those analyses, simulations and forecasts which are important to him. It will suffice if the program makes available pre-defined types of analysis, which are then calculated using those inputs which are relevant to the problem (for instance articles, key-accounts, time-periods, customer groups).

Such a program was created for the BrandSimulator and can be illustrated with some **examples** from the **fabric softeners data**. The data shown here is limited to analysis and simulation, since the forecast has already been shown above (see paragraph 3.4).

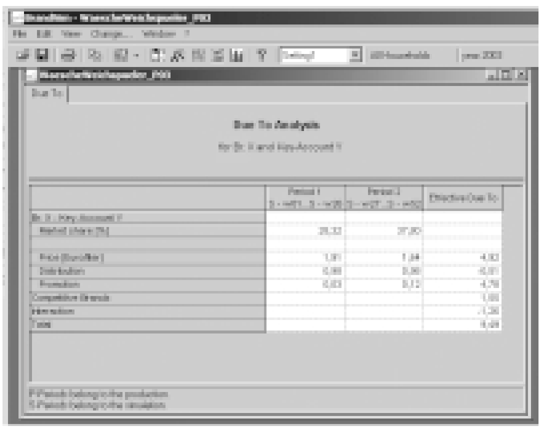
- Due-to-Analysis

One question that is often asked in marketing is what were the factors which caused a change in market share between one period (for instance the first half of the year) and another period (for instance the second half of the year). With the BrandSimulator it is possible to undertake very easily such a „Due-to-Analysis", which in the program goes through the following **stages**:

- For the first marketing mix variable relating to the brand to be examined, the values from the second period are copied over to the first period. The market share which was originally estimated by the model is then compared with the newly estimated market share. The difference can be explained by the marketing mix variable. In the example in *Figure 3,* the price of brand X in key-account Y falls from 1.91 € to 1.64 €. If the price from the first half of the year is replaced with the price from the second half (in another words falls already in the first half to an average of 1.64 €), that results in a market share which is 4.92 percentage points lower. The reduction in price has therefore generated a market share which was 4.92 **percentage points** higher.
- The price change is then made retrogressive, and the same exercize is repeated in turn with the other marketing mix variables.
- Then all the marketing mix variables for the competitive brands are changed together in order to determine the impact of competitive activities. This is also done retrogressive.

Figure 3

Example of a "Due-to-Analysis"



Explanation

In Key Account Y, Brand X won 9.49 percentage points market share in the second 6 months (weeks 27-52) compared with the first 6 months (weeks 1-26). 4.92 percentage points are due to the price reduction from 1.91 € to 1.64 €; -0.51 is due to distribution and 4.78 to promotion. Competitive brands had a positive influence (+1.55), while the interaction between the factors had a negative impact (-1.26).

Source: own calculations

- □ In the final stage the complete marketing mix of the client company and of its competitors in the first period is replaced by that from the second period. The comparison with the sum of single alterations can be regarded as an interaction effect.

Figure 3 shows that the increase in market share is primarily a result of the price reduction and increased promotional activity. Competitive activity and interaction only has limited importance.

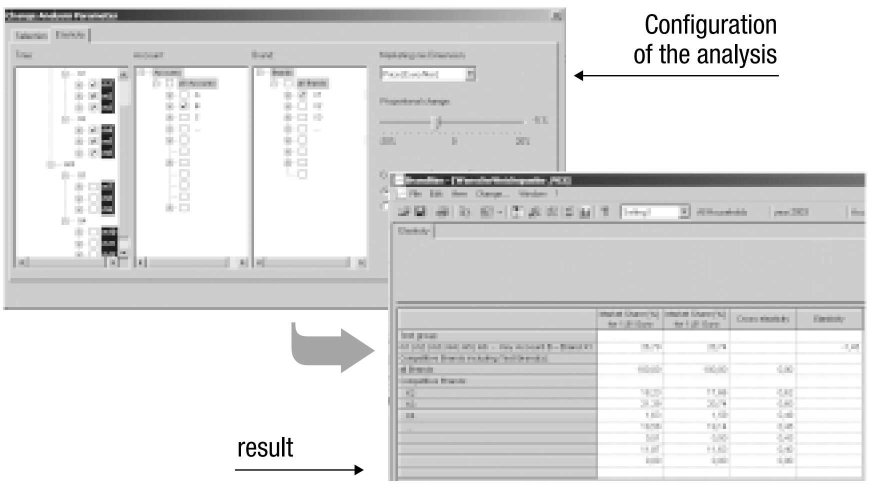
- ■ Price elasticity analysis

Price elasticity analysis forms a further category of analysis. To undertake this analysis, the following data is entered into the program (*Figure 4,* upper section of screen):

- □ The relevant period (here: the first six months of the year)
- □ The relevant key-account (here: key-account B)
- □ The relevant brand (here brand X1)
- □ The extent and direction of the price change being monitored. In this case, a price decrease of 5 % is being examined. Since prices are often immediately below a price

Figure 4

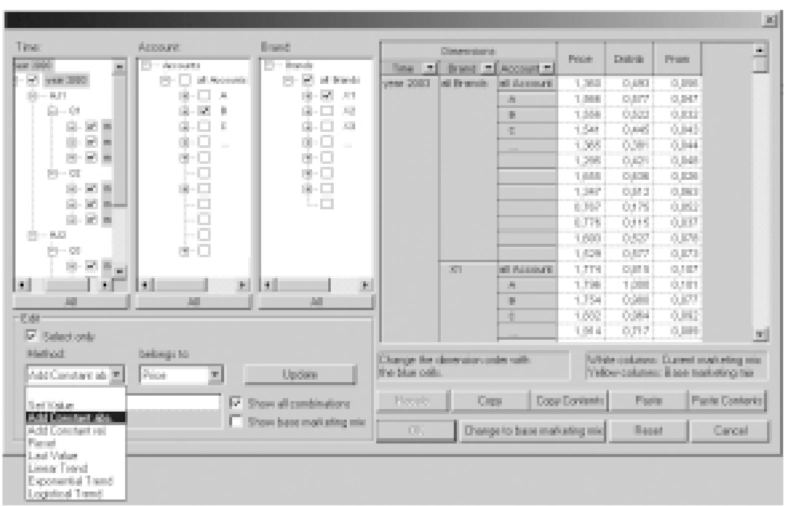
Example of an Elasticity Analysis



source: own calculations

Figure 5

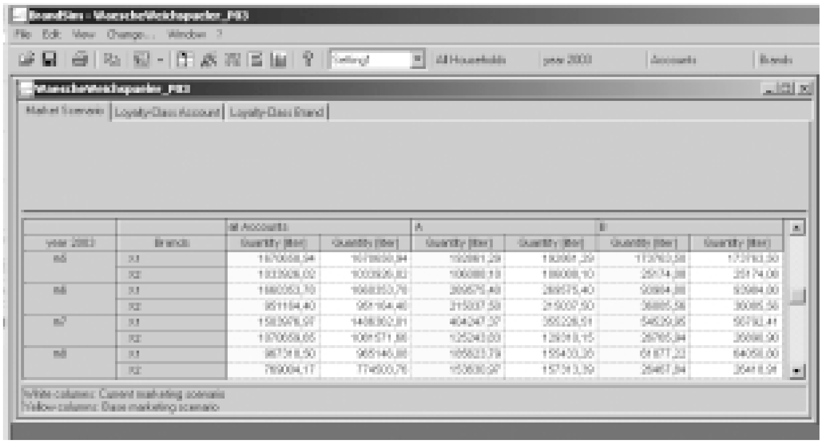
Simulation Data Input



Explanation: see text

Figure 6

Result Simulation



threshold, it is generally recommended here to enter price reductions. With price increases, there is a danger that price threshold effects and other price effects will become mixed up.

Figure 4 shows that brand X1, which is being considered has a price elasticity of -1.45, so that a price reduction of about 1% will result in a market share increase of about 1.45 % (not percentage points!). It also shows that this worked particularly to the disadvantage of Brand X2. Its cross-elasticity of 0.62 means that a price reduction of 1% by X1 tends to lead to a decline in market share for X2 of about 0.62 percent.

■ **Simulation**

Finally it is useful to describe a **Simulation**. *Figure 5* illustrates some of the many data entry possibilities: time-periods, brands, key-accounts and marketing mix variables can be freely selected and changed. Bottom left are several options: F.ex, how to change the price without typing all values manually. The price can be set to a certain value, can be increased/decreased by a fix amount or on percentual basis, or one of three trend models can be choosen to change the price. In this case, it is intended to check the effects of a 20 cent price increase for brand X1 in key-account A for months 7 to 12. *Figure 6* shows how the original sales (first column) change to the simulated sales (second column).

Figure 7

Case Study in the Beer Market

Question of the client:
What are the consequences for the total market, if the price for the four main brands goes down by 5%?

	Prices before (in Euro)	Prices after (in Euro)
Brand A	1.05	1.00
Brand B	0.93	0.89
Brand C	0.62	0.59
Brand D	0.82	0.78

	Market Share Volume before (in %)	Market Share Volume after (in %)	Price Elasticity
Brand A	5.0	5.3	- 1.23
Brand B	6.0	6.4	- 1.55
Brand C	4.5	4.6	- 0.41
Brand D	31.8	32.8	- 0.65

source: own calculations

Sales of brand X1 in key-account A decline significantly by 13%. But the figures for key account B show that part of the decline in sales is recaptured through sales of the other key-accounts. The columns for "all accounts" show that indeed a large part of the loss in sales in key-account A is compensated for. This shows a benefit of the Consumer Scanner Panel model, i.e. that migrations between key-accounts can be modelled.

Two case studies will point out how BrandSimulator results are used in practise:

In the first case study a client is responsible for a brand A in a food market. Besides A there are three other main brands in the same segment of the market: B and C and D. The client thinks about lowering the price for brand A. But this could result in a like-

Figure 8

Case Study in the Beverages Market

Question of the client:
What will be the impact on my Brand A and more generally on my brand portfolio (brands A and B), if I remove the size 50 cl of the Brand A?

	Market Share Volume before (in %)	**Market Share Volume after** (in %)
Brand A, size 50 cl	9.7	—
Brand A, size 75 cl	18.2	23.7
Brand B, size 50 cl	11.6	13.1
Brand B, size 75 cl	8.8	8.8
Brand B, size 75 cl promotional pack	7.4	7.8
Brand C, size 50 cl	3.6	4.5
Brand C, size 75 cl	11.6	11.6
Brand D, size 50 cl	10.6	11.2
Brand D, size 75 cl	7.7	7.8
Brand E, size 50 cl	10.9	11.6

source: own calculations

wise reaction of brands B, C and D. So the client wants to know the consequences for the total segment, if the prices for all four brands go down by 5 %.

A BrandSimulator model is run and prices of all four brands in all periods and all key accounts are lowered by 5%. All four brands are gaining market share. Price elasticities show that brand B is the most price sensitive with a value of -1.55. But also the clients' brand A is very price sensitive with a price elasticity of -1.23. The other two brands, brand C and D with price elasticities of -0.41 and -0.65 respectively are not very price sensitive. So brand A would profit more than average but less than brand B from the general price decrease (*Figure* 7).

In the second case study the client is responsible for the two beverage brands A and B in a market with a total of five brands. Most brands are offered in different pack sizes: 50 cl and 75 cl. The client thinks about stopping of the production of the 50 cl pack size of brand A. So he wants to know what the consequences are for brand A and brand B.

Figure 8 shows the market shares of all brands before and after the withdrawal of the 50 cl pack size of Brand A. Sales of the other brands in the 75 cl pack size are nearly unchanged. Most of the former buyers of brand A 50 cl switch to other products of brand A or to other products with the same size, 50 cl.

The size 75 cl pack size of brand A gains 5.5% in market share but this does not compensate for the loss of 9.7% for the 50 cl pack size. Brand B, which has three pack sizes, gains 1.9% market share, so even in the bigger context with brands A and B, the remaining brands of the client could not compensate for the loss. So the client has to see whether the cost reduction makes up for the loss in market share.

5. *Limitations and possible further developments in the Model*

In addition to the possibilities that have been described, it is appropriate to describe some of the **limitations** of the BrandSimulator.

One limitation arises from the fact that the whole marketing mix cannot be taken into account. Thus, poster, radio and cinema advertising are excluded as well as discontinuities in market structure, caused through either new products or through the repositioning of products or of retailing groups. Direct mail is in principle possible to be included but is not tested so far. Finally, the BrandSimulator, like every other model, assumes that the inter-relationships between marketing inputs and market shares remain stable.

A further limitation arises from the fact that consumer scanner panel data are necessary which only exist for packaged goods. Furthermore, a minimum number of purchase acts per purchaser are required. It is therefore not appropriate to model very infrequently bought products.

Finally a limitation arises through the fact that in the underlying logistical model, acceptance of the IIA assumption (4) is implied. As a result the substitution effects between products are evened out. This is mitigated as a result of every consumer having his individual relevant set and there is no possibility of substitution with a product which is outside the relevant set. Nevertheless this disadvantage exists.

Getting rid of the IIA assumption is the first possible **further development**. Attempts have already been made to do this, but up to now they have not been successful.

Two other possible further developments relate to the choice of shop. Since every household and every shop can now be located by GPS (5), the time required to get from home to the shop (on foot and by car) is known. This information can be of use just as much as the information whether a household buys anything from a certain shop, even if not in the category currently being investigated.

The BrandSimulator has already proved itself to be a model which provides more opportunities than all other previously known models which work on the basis of household panel data, but one where there are opportunities for further important improvements.

Notes

(1) EAN stands for European Article Number and refers to the number contained in the barcode, which is printed on almost all industrially packed goods which are in regular daily use. On EAN see Günther et. al. 1998, p. 120.

(2) This solution was originally put forward by Prof. Daniel Klapper, University of Kiel.

(3) TV-viewing behaviour and buying behaviour are collected simultaneously in part of A. C. Nielsen's Household Panel (*Turgeon/Löwenbein* 1993; *Heuer/Hirvonen* 1993; *Griese* 1993; *Wildner* 1994). This involves though problems in terms of willingness to co-operate on the part of panel members (*Anonymous* 1995; *Wildner* 2000). As a result, in the GfK Household Panel only data on purchasing behaviour is collected, while in the AGF/GfK Television Audience Research Panel only data on viewing behaviour is collected.

4) IIA stands for „independence of irrelevant alternatives". This reflects the fact that with a Logit Model, changes in the market share of a product (for instance due to a price increase) require at the level of the individual, a compensating change in the market share of the other products in comparison with their market share before the change. The following fictitious example makes clear that this assumption is certainly not in general terms correct: a consumer buys coffee brands A and B which are caffeine-rich for breakfast drinking, and coffee brands C and D, which are caffeine-free for drinking in the afternoon. All four coffees are bought in equal volumes. Let us assume that if brand D is withdrawn from the market, the consumer will primarily buy increased volumes of brand C. But in a Logit-Model, A, B and C would benefit to an equal amount from the withdrawal of D (for more on the IIA assumption, see *Train* 2003, pp. 54 ff.).

5) GPS-technology (the abbreviation stands for „Global Positioning System"), has since 8.12.1993 enabled locations to be determined with the aid of satellite technology to within about 10 meters. To make use of satellite technology, all the addresses and the complete road network in Germany (as also in other countries) were located. As a result, it is possible to work out the journey-time by car or on foot between all addresses in Germany (see http://www.kowoma.de/gps/Geschichte.htm).

References:

Ailawadi, K. /Neslin, S. (1998): The Effect of Promotion on Consumption: Buying More and Consuming It Faster, in Journal of Marketing Research, Vol. 35, pp. 390–398.

Ailawadi, K. /Gedenk, K. /Lutzky, C. /Neslin, S. (2005): The Benefits of Promotion-Induced Stockpiling, unpublished working paper.

Amstutz, A. (1967): Computer Simulation of Competitive Market Response, Cambridge.

Anonymous (1995): Single Source – ein Methodentest der AGF/GfK Fernsehforschung [A Methodological Test by AGF/GfK Television Research], published by GfK Fernsehforschung, Nürnberg.

Author and year unknown: www.kowoma.de/gps/Geschichte.htm

Fader, P. /Hardie, B. (1996): Modelling Consumer Choice Among SKUs, in: Journal of Marketing Research, Vol. 33, pp. 442–452.

Günther, M. / Vossebein, U. / Wildner, R. (1998): Marktforschung mit Panels - Arten - Erhebung - Analyse - Anwendung [Panel Market Research – Types – Data Collection – Analysis – Applications], Wiesbaden.

Griese, U. (1993): Single Source: Neue Möglichkeiten für Mediaplanung und Analyse [New Possibilities in Media Planning and Analysis], in: planung & analyse 6/1993, pp. 58–62.

Gupta, S. (1988): Impact of Sales Promotions on When, What and How much to buy, in Journal of Marketing Research, Vol. 25, pp. 342–355.

Heuer, H.-D. /Hirvonen, P. (1993): Wie wirken Promotions? Neue Wege der Evaluierung mit dem Nielsen Single Source Panel [How do Promotions Work? New Types of Evaluation with Nielsen Single Source Panel], in: planung & analyse 4/1993. pp. 23–27.

Klenger, F./Krautter, J. (1973): Simulation des Käuferverhaltens, Teil 1: Werbewirkung und Käuferverhalten, Teil 2: Analyse der Kaufprozesses, Teil 3: Computermodell des Käuferverhaltens [Simulation of Purchase Behaviour, Part 1: The Effect of Advertising and Purchase Behaviour, Part 2: Analysis of the Buying Process, Part 3 Computer Model of Buying Behaviour], Wiesbaden.

Klapper, D. /Ebling, C. (2004): GfK-Projekt: Analyse von Panel-Daten mit Markenwahlmodellen [GfK Project: Analysis of Panel Data with Brand Choice Models, unpublished working paper], unveröffentlichtes Arbeitspapier.

Kopperschmidt, K. (2005): Statistische Analyse von Kaufverhaltensmodellen [Statistical Analysis of Purchase Behaviour Models], Gießen.

Lavington, M.R. (1972): Ein Mikrosimulationsmodell der Nachfragereaktion beim Konsumgütermarketing [A Micro-simulation of Reactions in Terms of Demand in Consumer Goods Marketing], in: *Kroeber-Riel, W.* (Ed.): Marketing theory, pp. 332–358, Köln.

Little, J. (1974): Brandaid: A Marketing Mix Model, Part 1: Structure, Part 2: Implementation, Calibration and Case Study, in: Operations Research 23, pp. 628 ff.

Roberts, A. (1994): Media exposure and consumer purchasing: an improved data fusion technique, in: Marketing and Research Today, August 1994.

Schneeweiß, H. (1974): Ökonometrie [Econometrics], Würzburg.

Train, K. (2003): Discrete Choice Methods with Simulation, Cambridge.

Turgeon R./Löwenbein O. (1993): Zielgruppenbestimmung: Neue Wege auf der Basis von Nielsen Single Source [Target Group Identification: New Methods on the Basis of Nielsen Single Source], in: planung & analyse 3/1993, pp. 23–27.

Vossebein, U./Wildner, R. (1992): Komplexe Fragestellungen erfordern komplexe Lösungsansätze: Ein Marketing-Mix-Modell im Praxistest [Complex Problems Require Complex Solutions: A Marketing Mix Model on Practical Test], in: planung & analyse 4/1992, pp. 56–61.

Wildner, R. (1990): Simulation and Prognosis: New Ways for Getting More Information out of Integrated Panel Data, in: EMAC / ESOMAR Symposium on "New Ways in Marketing and Marketing Research", Athens, 24.–27. October 1990, Amsterdam.

Wildner, R. (1991): Nutzung integrierter Paneldaten für Simulation und Prognose [The Usage of Integrated Panel Data for Simulation and Forecasting], in: Jahrbuch der Absatz- und Verbrauchsforschung 2/1991, pp. 114–130.

Wildner, R. (1994): Die Grenzen von Single Source [The Limitations of Single Source], in: planung & analyse 1/1994, pp. 22–27.

Wildner, R. (2000): Messung von Werbewirkung mit fusionierten Paneldaten [Measurement of Advertising Effectiveness with Merged Panel Data], in: Jahrbuch der Absatz- und Verbrauchsforschung 3/2000, pp. 242–260.

Wildner, R./Kindelmann, K. (1997): How to measure and judge TV ads' effectiveness in single source data, in: Revolutions in Panel Research, published by ESOMAR, Amsterdam.