# Statistical modelling in market research

Anatoly Zhigljavsky,
School of Mathematics, Cardiff University, Cardiff, U.K.

Mathematical modelling is a key element of quantitative marketing and helps companies around the globe in making important marketing decisions about launching new products and managing existing ones. Most mathematical models used in marketing research are either purely statistical or include elements of statistical models.

An extensive discussion (by the top market research academics) of the state-of-art in the field of marketing modelling and its prospects for the future is contained in [1], a special issue of the International Journal of Research in Marketing. One can consult [1] for many references related to the subject; see also recent books [2-5].

We look at the field of market modelling from a view-point of a professional statistician with twenty years of experience on designing and using statistical models in market research. We start with distinguishing the following types of statistical models used in market research:

(1) Direct simulation models;

(2) Standard statistical models;

(3) Models of consumer purchase behaviour;

(4) Dynamic models for modelling competition, pricing and advertising strategies;

(5) Statistical components of inventory and other management science models.

Let us briefly consider these types of models separately.

*1. Direct simulation models.* These are specialized models based on attempts to directly imitate the market (for example, via the behaviour of individual customers) using a synergy of stochastic and deterministic rules. These models were popular 20-30 years ago but are less popular now. The reasons are the lack of predictive power, huge number of parameters in the models and impossibility of their validation.

*2. Standard statistical models.* All standard statistical models and methods can be used in market research, see [4-7]. Most commonly, the following statistical models are used:
– various types of regression,
– ARIMA and other time series models,
– Bayesian models,
– models and methods of multivariate statistics; especially, structural equation and multinomial response models, conjoint, factor and principal component analyses.

*3. Models of consumer purchase behaviour.* Several types of statistical models are used for modelling consumer purchase behaviour including brand choice. The following three basic models (and some of their extensions) have proved to be the most useful: Mixed Poisson processes, the Dirichlet model and Markovian models.

The mixed Poisson process model assumes that a customer makes his/her purchase according to a Poisson process with some intensity $\lambda$ where $\lambda$ is random across the population. In the most popular model, called Gamma-Poisson, $\lambda$ has Gamma distribution (with two unknown parameters); this yields that the number of purchases for a given period is the Negative Binomial Distribution. Typical questions, which the Poisson process model answers, is the forecasting of the behaviour of the market research measures (like penetration, purchase frequency and repeat buying measures) in the form of the so-called growth curves. Extensions of the mixed Poisson models cover the issues like the zero-buyer problem (some zero-buyers do have a positive propensity to buy but some other don't), seasonality of the market and the panel flow-through.

The Dirichlet model is a brand-choice model. It assumes that customers make their brand choice independently with certain propensities; these propensities are different for all customers and are independent realizations from the Dirichlet distribution which parameters are determined by the market shares of the brands. In Markovian brand-choice models, the propensity to buy a given brand for a random customer may vary depending on either the previous purchase or other market variables. These models are more complicated than the mixed Poisson process and Dirichlet models but in some circumstances are easily applicable and sometimes are able to accurately describe some features of the market.

Of course, the models above are unrealistic on the individual level (e.g. few people have the Poisson process pattern as their purchase sequence). However, these models (and especially the mixed Poisson model) often fit data extremely accurately on the aggregated level (when the time period considered and the number of customers are sufficiently large). These models can be classified as descriptive (rather than 'prescriptive') and help in explaining different aspects of market research dynamics and some phenomena related to the brand-choice.

*4. Dynamic models for modelling competition, pricing and advertising strategies.* There is extensive literature on this subject, see e.g. [8]. The majority of the models are so-called differential games or simpler models still written in terms of differential equations. The models are deterministic and the statistical aspect only arrives through the assumption that the data contain random errors. Statistical modelling part is therefore negligible in these models. Alternatively, in some Markovian brand-choice models mentioned above, there is a option of inclusion market variables (e.g. promotion) into the updating rule for the buying propensities. These models are proper stochastic models but they are often too complicated (have too many parameters) and therefore difficult to validate.

*5. Statistical components of inventory and other management science models.* Inventory and other management science models applied in market research are typically standard models of Operations Research, see [9] for a recent review of these models. Despite these models often have a large stochastic component, they do not represent anything special from the statistics view-point.

Statistical models are used for the following purposes: (a) forecasting the market behaviour of a new brand to prepare its launch; (b) managing existing brands. In case (a), the models are usually based solely on standard statistical models, type 2 above. Sometimes, other types of models (especially, large simulation models, type 1) are used too. A lot of specific market research data are often collected to feed these models. These data includes market surveys, various types of questionnaires and focus group research in direct contact with customers. All available market data, for example economic trends and specific industry sector reports, is used too. In case (b), the models are used for making decisions about pricing, promotion and advertising strategies, production and inventory management etc. All available statistical models and methods are used to help managers to make their decisions.

While reading academic papers and books on marketing research, one can get an impression that mathematical and statistical modelling in marketing is a mature subject with many models developed and used constantly for helping market research managers in working out their decisions. Indeed, there are many models available (some of them are quite sophisticated). However, only a small number of them are really used in practice: the majority of practical models can be reduced either to a simple regression or sometimes to another standard model among those mentioned above. One of the reasons for this gloomy observation is the fact that managers rarely want a description of the market. Instead, they want 'a prescription'; that is, a number (with a hope that no confidence interval is attached to this number) which would lead them to a right decision. Another reason is the fact that only a very few models used in market research satisfy the following natural requirements for a good statistical model: (a) simplicity, (b) robustness to the deviations from the model assumptions, (c) clear range of applicability, and (d) empirical character, which means that the models have to be built with the data (and data analysis) in view and with the purpose of explaining/fitting/forecasting relevant data.

Despite huge amounts of market data is available to analysts, these data are typically messy, not reliable, badly structured and become outdated very quickly. Development of reliable statistical models dealing with such data is hard. The progress in understanding all these issues and tackling them by means of the development of appropriate models and making them correctly applicable is visible but it is justifiably slow.

**References**

[1] International Journal of Research in Marketing (2000), v. 17, No. 2-3 (Special issue on Marketing Modelling on the Threshold of the 21st Century)

[2] Wierenga B., ed. (2008) Handbook of Marketing Decision Models, Springer.

[3] Wittink D.R., Leeflang P.S.H. and Wedel M., Naert P.A. (2000) Building Models for Marketing Decisions, Kluwer Academic Publishers, Boston, MA.

[4] Mort D. (2001) Understanding statistics and market research data, Europa publications, London.

[5] Zikmund, W.G., Babin B.J. (2009) Exploring marketing research, South-Western Educational Publishing, Florence KY.

[6] Rossi P.E., Allenby G.M., McCullog R. (2005) Bayesian statistics and marketing, Wiley-Blackwell, N.Y.

[7] Hanssens D.M., Parsons L.J., Schultz R.L. (2003) Market response models: econometric and time series analysis, Springer.

[8] Erickson G.M. (1991) Dynamic models of advertising competition: open- and closed-loop extensions. Springer.

[9] Ingene C.A., Parry M.E. (2004) Mathematical models of distribution channels, Springer.