Market modelling · 1
The basics of modelling

Chuck Chakrapani

Marketing research as a snapshot in time
Marketing research findings are like photographs. They are snapshots in time. We can extend this analogy further. The quality of a photograph depends on the skill of the photographer and the equipment used in capturing the moment; the quality of research will depends on the survey instrument, sampling and statistical procedures used and the skill of the researcher. A photograph can distort reality by putting what should be in focus out-of -focus and what should be out-of-focus in-focus; so can research. The technical excellence of a photographs relatively easy to judge. But it is not easy to judge how truthful a representation the photograph is of the content represented. This is true of research as well. A technically good research project may not represent reality any more than a technically good photograph.

Yet, like a photograph, research can be used as a valuable tool to represent reality. This, of course, requires knowledge of the factors that distort true representation.

"The MOST for the LEAST"
Business operates by maximizing the return on investment, including what is spent on research. However, there seems to be an increased tendency to define the return in quantitative rather than in qualitative terms. The constant quest for "the MOST for the LEAST" (as a telecommunications company ad has put it) has resulted in overly long questionnaires with convoluted questions, misuse of qualitative research (often using it as a substitute for quantitative research), dubious sampling procedures, accepting the most easily available respondents as opposed to respondents who should be contacted and other less than sound research practices. As a result, it would seem that we have a large body of research information that is of dubious value generated by research practitioners and buyers following 'the MOST for the LEAST' philosophy.

From the present to the future
Research information, although a snapshot in time, has an additional function. It serves as input to future courses of action. For instance, the usefulness of the research data ('snapshot in time') depends on whether we can use it to market our product better - increase sales, reach the customer more cost effectively, increase customer loyalty or whatever. If we cannot project the data or their implications into the future, the research is fossilised and is of limited value. A good photograph is appreciated because it is pleasing to look at, but, in research, such aesthetic considerations are replaced by more mundane utilitarian considerations.

The utilitarian aspect of research is the beginning of model building. Many researchers have chosen to specialize in the 'snapshot aspects' of research leaving the details of the use of research to marketers, modellers and statisticians. While it is quite legitimate (probably even desirable) for modelling and research to be separate disciplines, this practice has created problems over the years. Researchers' lack of familiarity with the limitations of a model coupled with the modeller's lack of familiarity with the implications of research data have resulted in models that are inappropriate.

Not surprisingly, although researchers (with a little help from statisticians) have been busy developing models, use of these models have been limited and inconsistent. Despite a proliferation of modellers hawking their wares, systematic model building has not been evident. That is, until recently.

Two new forces
Of late there has been a renewed interest in modelling. I believe that there two are major reasons for this renewed interest.

First, the universal availability of personal computers have made it easy for anyone to manipulate a model or at least a part of it. For example, if there is a forecasting model, a marketer can simply change a few numbers in his/her spreadsheet and study the effects of 'what-if' scenarios. Such ease of manipulation enables the user to assess the reasonableness of a model under a wide variety of conditions. Although the user (or the researcher) is not the modeller, the ease of use has encouraged greater communication between the modeller, the researcher and the marketer.

Second, there has been an explosion of access to information which is both extensive and inexpensive. This is obviously related to the near-universal availability of personal computers and the growth of information technology. For instance, if we carry out a large scale market survey, we may be able to append other information (such as geodemographic characteristics) to the survey. Practically a limitless amount of information can be added to a database. While some of this information is of questionable value, the availability of such information itself has resulted in an increased curiosity about its use.

Many companies have started building databases to enable them to precisely target potential customers. The emergence of database marketing as a major force has placed new tools in the hands of the marketer.

There is no easy way of handling such a large quantity of information short of using it as input to a model that will enable us to understand the market better.

What is a model?

A model is someone's idea of how things work. A mathematical model is someone's idea of how things work, expressed in terms of mathematical notations and equations. Obviously if someone's idea of how things work is wrong, so is the model.

There is a widespread misconception that quantitative models are driven by underlying mathematical or statistical principles. This has lead to the unquestioning acceptance of many mathematical models without critically examining the underlying assumptions. It is best to remember that what drives a model is the thinking behind it, and not the arcane mathematical form in which it is expressed. If a modeller cannot explain in plain English what s/he is doing, then perhaps the modeller does know what s/he is doing. Modelling is a process of clarifying our thinking, not a way of obscuring it. Sound quantitative principles may help us translate out thinking into quantifiable formulas, but quantified formulas do not imply sophisticated thinking.

Let us consider a simple mathematical model:

$S_n = C * S_{n-1}$
Where,
S_n = Sales of a product on year n
S_{n-1} = Sales of a product in the previous year ($n-1$)
C = Constant increase or decrease per year

The model simply states that the sales will go up (or down) each year by a constant percentage. Where do we get the constant percentage from? From the past data or from guesswork.

When a model such as this one is presented, it may look as though the mathematics behind the model enable us to estimate next year's sales. But, in fact, mathematics has nothing to do with the estimates. The critical assumptions are made by the modeller and not by the underlying mathematics. Even in a model that is as simple as the one mentioned above, there are many critical assumptions:

1. Sales in any give year is dependent on the sales achieved in the previous year.

2. The annual increase is a constant percentage, year after year.

3. The data on which we base our constant on is reliable and valid.

All the above assumptions can be called into question. If they are not true, they can invalidate the model. Mathematics did not create these assumptions, the modeller did. Mathematics simply formalized the assumptions.

While a marketer or a researcher may not know how to set up and operate a mathematical model, they can (and should) provide input to the model. The above example clearly illustrates that it is the assumptions made by the modeller and not the mathematics that was primarily responsible for the results generated by the model.

Models are generally based on past data. To the extent that past data are not reliable, we will build models that are error prone or even highly misleading. It is in this context that we should view the pitfalls of the data generated by the 'most for the least' approach.

Modelling is too serious an exercise to be left to professional model builders.

Prerequisites of a model
There are three prerequisites of a good model.

1. The model's assumptions should be valid, or at least reasonable.

2. The data on which the model will be developed should be valid and reliable.

3. The mathematics behind the model should be sound.

The first two aspects of model building should be the joint responsibility of the marketer/researcher and the model builder. Competence in model building does not imply competence in the other two areas.

Five steps to building a model
What then constitutes a model? A basic marketing model would be built using the following process:

1. The modeller starts with a question that relates to the real world.

What is the relationship between the price of my product and customers' demographic characteristics?

2. The modeller then identifies the possible nature of the relationship.

I can sell more of my product to those who have characteristics X,Y and Z but less to those with characteristics R, S and T.

3. The modeller translates the relationship into mathematical notations and equations.

Sales = f (X, Y, Z, R, S, T)

(This step assumes that the model is quantitative. Most formal marketing models are quantitative.)

4. The modeller then collects or retrieves data to apply the model.

Data needed: Sales of the product, quantitative information on characteristics X, Y, Z, R, S and T.

5. The model is then developed tested, retested and refined.

Pitfalls, assumptions and questions
The process of model building may look simple, but there are a number of pitfalls, implied assumptions and questions. We can review this step by step.

FIVE STEPS TO BUILDING A MODEL

1. Start with a question about the real world

2. Identify the possible nature of the relationship

3. Express (2) above as a mathematical relationship

4. Retrieve or collect data to confirm and/or refine the model

5. Test, refine and retest the model

Step 1. The modeller starts with a question that relates to the real world. Let us start with an apparently simple question such as

What is the relationship between the price of my product and the purchase behaviour of customers' with different demographic characteristics?

An experienced modeller may not find the question as simple as it apparently looks.

Let us start with the price of the product - Are we talking about the absolute price or the price relative to competitors? Should the model include or exclude special deals? Are we talking about the price we offer the distributors or the price to the end-user? Are we describing the relationship between price and purchasing behaviour in the short term or in the long run?

As we move from the price to purchase behaviour, we are faced with more questions: Is the product distribution comparable in different parts of the country? Is there less competition in the East than in the West? Are there substitute products available in one region of the country as opposed to another? Are there extraneous variables affecting purchase behaviour which could mistakenly be attributed to price? What are the factors that could confound the results? For example, if the consumption of a product is on the decline and independently of this decline, the price is on the rise, could we not mistakenly conclude that price rise affects purchase behaviour? What other factors can potentially confound the model?

What demographic characteristics should we include in the analysis? Do we have information on all demographic variables?

Step 2. The modeller then identifies the possible nature of the relationship.

I can sell more of my product to those who have characteristics X,Y and Z but less to those with characteristics R, S and T.

What do we mean by 'more'? More in terms of total volume or per capita units? Is 'more' large enough to be meaningful? Is the model even worth building if there is evidence that the cost of marketing to people with characteristics X, Y and Z will be far too high?

Step 3. The modeller translates the model into mathematical notations and equations.

Sales = f (X, Y, Z, R, S, T)

Given the same assumptions, a modeller may consider many alternative mathematical models. They may be based on different mathematical premises and vary in their level of sophistication. For this step we need a highly competent model builder with sound knowledge of the underlying mathematical/statistical principles involved in model building.

Step 4. The modeller then collects or retrieves data to apply to the model.

Data needed: Sales of the product, quantitative information on characteristics X, Y, Z, R, S and T.

As we noted earlier, the quality of data can affect the model. A model that is based on unreliable or invalid data can provide misleading results.

Step 5. The model is then developed tested, retested and refined. Here again we need to depend on the model builder's competence and knowledge.

Models are here to stay

It seems fairly obvious that the use of modelling will increase in the future. Inexpensive computer power, increased availability of databases and increasing sophistication in the area of data base marketing and other related trends will ensure that.

GIGO

I sometimes fear that, in the rush to make databases more manageable, in the rush to produce 'sophisticated' models, the quality of the data and assumptions on which the models are based may be overlooked.

The GIGO principle (Garbage In and Garbage Out) might still apply, but the acronym may come to stand for something else: *Garbage In and Gospel Out.*

In any case, as researchers we cannot afford to be ignorant of what has been happening in the area of modelling. In subsequent articles in this series we will look closely at some models, especially the 'predictive' ones.

Dr Chuck Chakrapani is a Toronto-based consultant, author and speaker who works internationally.