Market modelling · 8
Structural Equation Models - 2

Chuck Chakrapani

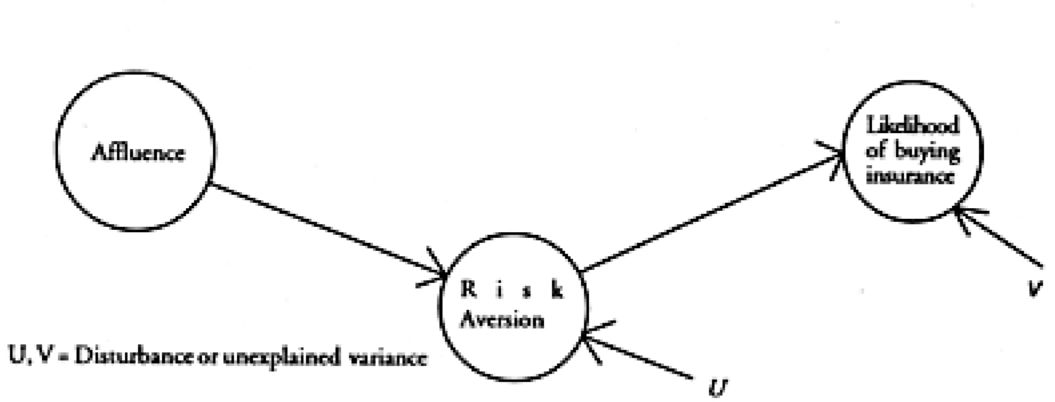
Identifying the latent variables
In an earlier article I suggested that Structural Equation models deal with latent variables, ie. variables that cannot be directly measured but have to be elicited from other measured variables. What procedure can we use for such elicitation? Obviously the procedure should be a standard one. The most widely used procedure is factor analysis. Factor analysis enables us to (mathematically) combine different yet related attributes into a smaller number of 'factors'. It is a process of assigning different weights to different attributes. The weights so assigned are expected to reveal the similarity of groups of attributes pointing to some underlying constructs. For instance, if attributes such as 'unbreakable', 'lasts a long time', 'made well' and 'sturdy' are correlated, factor analysis will assign weights to these attributes and present them as a single factor. Looking at the attributes that contribute to this factor, we may infer the latent variable: Durability.

Known and unknown parameters
The first objective is to specify the model so it can be tested. A model is specified by the marketer based on past knowledge and theoretical expectations. SEM contains implied structural equations and specified causal paths among latent variables. Some model parameters (such as correlations among latent variables) are known. Other parameters (such as the correlation between exogenous variables and paths between variables) are unknown. In specifying the model, the modeller should bear in mind that the known parameters should equal or exceed unknown parameters. When the number of known parameters equal the number of unknown parameters the model is known as *just-identified.* When known parameters exceed the unknown parameters, the model is *over-identified* ; and when unknown parameters exceed known parameters, the model is under-identified. Unknown parameters can be estimated for over-identified and just identified models but not for under-identified models.

Calculating the known and the unknown parameters
The number of known parameters is calculated by the formula N(N - 1) /2, where N = the total number of variables in the model (both exogenous and endogenous). The number of unknown parameters in the model is the sum of all paths, correlations and correlated disturbances among variables.

SEM: An example
As an example consider the following model: Affluence leads to Risk Aversion which in turn leads to the buying of insurance products. Here, Affluence, Risk Aversion and Likelihood of Investing in an Insurance Product are all latent variables (see Exhibit 1).

Exhibit 1: Hypothetical SEM of Insurance Purchase Behaviour

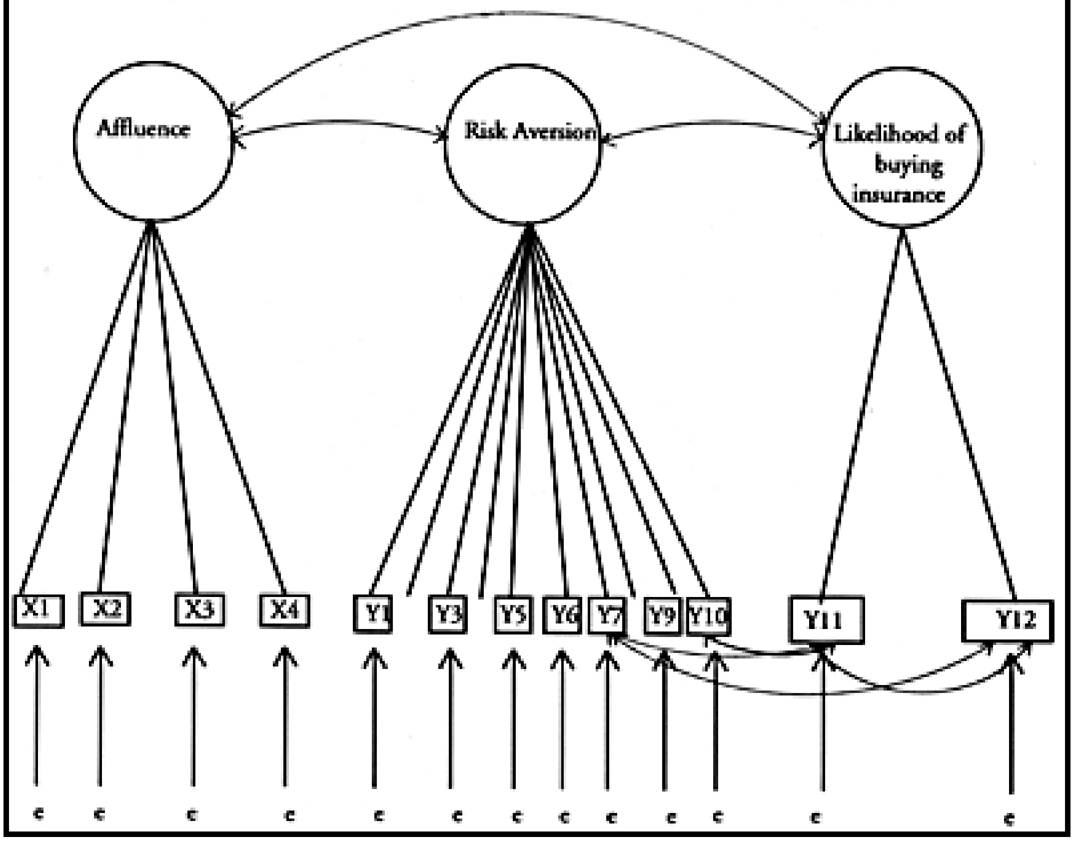


Specification
1. The exogenous variable Affluence is measured by four observed indicators: the person's net worth, current income from all sources, home ownership and debt level (all observed variables). Risk Aversion is measured by the respondent's answers to a series of 10 questions that deal with attitudes to risk. Finally, the Likelihood of Investing an Insurance Product is measured by two observed variables: the reported probability of buying each type of insurance in the next 12 months and the reported amount invested in each of them.

2. Then correlated errors are specified (on the basis of measurement or an hypothesized relationship). For instance, if a person rates herself as risk averse in her daily life and also reports how likely she is to buy insurance in the next 12 months, the errors attached to these two variables are likely to be correlated.

An example of the specified model is given in Exhibit 2.

Exhibit 2: Hypothetical Measurement Model of Insurance Purchase Behaviour



Model testing using LISREL
Model testing is carried out by running the data through a SEM program. In the last issue of Imprints, I mentioned several programs capable of handling SEM. Any one of the programs mentioned (or any other that you may be aware of) can be used to estimate the model parameters.

Suppose we apply LISREL to the above problem to estimate the parameters. We should be able to obtain the following information from the output:

1. Lambda Matrix. This matrix shows how well the observed variables correlate with ('load on') the latent variables. This information is important since, unless there is a strong relationship between observed and latent variables, the model is unlikely to be robust.

We can examine this matrix in the same way we would a factor loading matrix when we do factor analysis. We look to see if the loadings are large enough. Although there are no hard and fast rules here, we would expect to see loadings that exceed .5, preferably .7. (A Lambda t-value matrix provides the statistical significance of these values.)

2. Psi Matrix. This matrix shows how well latent variables correlate among themselves.

In our example, what is the correlation between the latent variables Risk Aversion and Likelihood of Buying Insurance? Suppose that the correlation between the two variables is .7. It is clear that these two latent variables are indeed correlated in the way that we had hypothesized. (A Psi t-value matrix provides the statistical significance of these values.)

3. Theta matrix. This matrix provides the estimates of the correlated measurement error specified in the measurement model.

In our example, we hypothesized that there is a correlation between an observed variable related to risk aversion and another related to the likelihood of buying insurance. The obtained correlation between the two variables was .57. This is large as well as statistically significant. This result simply means that this hypothesis is now confirmed and the relationship can be included in the final model.

In addition to the above, LISREL provides standardized residuals for each observed indicator. These values are a measure of systematic error.

When we look through the residuals in our example, we find that most of the residuals fall below ±2 standard deviations (not shown here). This means that there are no outliers that might distort the model.

After estimating individual parameters of the model, LISREL then subjects it to a Maximum Likelihood Chi-squared analysis to identify the overall fit of the model.

In our example, the chi-squared value turned out to be statistically significant, meaning that our model is confirmed by the data to which it was applied.

If the overall fit were not significant, it would obviously mean that the model is not supported by the data.

When the model has a good fit, we may still have some doubts as to its applicability to other sets of data. However, when the model has a poor fit, we have no alternative but to reject or revise our model. An experienced model builder obviously would not fully develop and test a model without some preliminary or intermediate indications that it is a reasonable model of the underlying process.

Improving the model

LISREL also provides suggestions (called the modification indices) throughout the modelling processes. They refer to what changes need to be made to the model to improve the fit. These suggestions, if accepted, are likely to result in a model that fits the data better.

However, one should be careful. These suggested improvements are based solely on statistical improvement to the model and have little to do with the logic of it. Therefore, when considering the suggestions, we should make sure that they are reasonable and they do not alter (in an unreasonable way) our basic ideas about the nature of relationships.

Causal models in marketing

How useful is causal modelling in marketing? We don't know yet, because SEM is widely discussed but not widely used. Since SEM uses both observed and latent variables, the ultimate relationship between latent variables can be misleading, if the observed variables have only a weak (albeit statistically significant) relationship with the latent variables.

Nevertheless, SEM provides a formal basis and a means of testing for many marketing assumptions. This alone makes it worth considering.

Book Review
The New Positioning

by Jack Trout with Steve Rivkin
Published by McGraw-Hill 175 pp. $31.95+GST

Jack Trout, who claims to have coined the word positioning in an article published in Industrial Marketing 25 years ago, is the well-known co-author of the book Positioning: The Battle for Your Mind. His current book consolidates and extends his earlier observations on the topic. The book's core axiom is that the human mind cannot cope with the complexity around it. Unless we understand this, we cannot win the battle for the mind (and impart marketing messages). There are five corollaries to this central axiom.

1. Minds are limited. They can take in only 7 chunks of information at a time. They filter and take in only information in which they are interested.

2. Minds hate confusion. Minds go after simplification. So oversimplify your message.

3. Minds are insecure. People try to avoid risks, be they monetary, functional, physical, social or psychological. A would-be persuader should be able to appeal to emotions, use appeals to suggest that 'everyone is doing it', use testimonials, use the bandwagon effect and invoke heritage.

4. Minds don't change. So you might be better to reclaim old ideas: like KFC bringing in another actor to play Colonel Harland Sanders in its ads (although he died several years ago).

5. Minds can lose focus. This corollary has a startling implication: line extensions can weaken the brand image.

The second part of the book deals with issues of repositioning. It discusses repositioning of companies from four different sectors.

The third part of The New Positioning deals with a number of topics that are of interest to the marketer: naming a category, using PR for positioning, and positioning pitfalls.

The author follows his own prescription. This book is short, simplified, easy to read and uses a large typeface. It can be read in one evening. Although many of his comments make eminent sense, how he arrived at them is not always clear-the author does not consider alternative explanations of the facts presented and the book does not contain any bibliography or source material. Because this book is interesting and useful, it is disappointing to note that it is written without due care in some places. For example the author describes the decidedly European Edward de Bono (born in Malta, educated and taught in England and domiciled in Italy) as "one of America's best thinkers". We don't need that much oversimplification!

All the same, this book provides stimulating material on positioning as well as on the human mind in general. Recommended read.

Chuck Chakrapani

Dr. Chuck Chakrapani of Standard Research Systems is a Toronto-based consultant, author and seminar leader. He works internationally.