Market modelling · 7
Structural Equation Models - 1

Chuck Chakrapani

Causal inferences with available data
In path analysis we were concerned with causal relationships among different marketing variables. Structural equation models extend this orientation. They permit us to test our hypotheses and make causal inferences about the effect of certain variables on other marketing variables without specifically designing a study to do so. The advantage is obvious. A marketer attempting to solve a marketing problem may have the advantage of solving a problem with the data he or she already has, rather than having to collect more data. Structural Equation Models (SEM) allow us to make causal inferences with correlation data rather than with experimentally manipulated data.

Latent variables
Our focus in these articles is on latent or unobserved marketing variables. We will discuss the basic conditions of causation along with the theory of SEM models using latent variables.

Conditions of causation
SEM consists of a series of statements that describe cause (x) and effects (y) of the form X -> Y. For instance, we can state that advertising (X) causes sales (Y). For a causal statement to be specified this way, three conditions need to be fulfilled. The first condition is *temporal ordering* . This means X must precede Y in time. In our example, sales that occur prior to advertising cannot be specified as the Y variable. The second condition is that the relationship should be *observed and measurable* . In our example, a logical positive relationship between advertising and sales in general can be demonstrated and can be measured. The third condition is *nonspuriousness*. This condition implies that the relationship between X and Y is genuine and cannot be attributed to other factors. A spurious relationship can exist, for instance, if an advertiser advertised only when sale prices were offered. If this were the case, the relationship between advertising and sales can be potentially spurious in the sense that the 'real' relationship could be between reduced prices and sales rather than advertising and sales.

Theoretical basis of the model
In addition to the conditions of causation listed above, a good SEM should also be based on sound theory. We can use observed relationships more confidently if there is a theoretical basis for them. A theoretical structure may help us minimize spurious specification of relationships on the basis of observed facts. As we have been discussing throughout this series of articles, even when there is a theoretical foundation, and all the three conditions of causation are met, causal statements are *never proven* by the model. They are simply either confirmed or unconfirmed. The validity of the confirmed relationships will depend on the adequacy of the theoretical basis of the model specified.

Structural equation models
An SEM is a causal model which hypothesizes causal relationships among marketing variables to explain a marketing phenomenon. For instance, we may want to determine why some customers buy an RRSP after being exposed to a direct mail brochure while others don't. A careful review of earlier observations and theoretical relationships may identify income level, age group and attitude towards retirement as the main causal factors. We can then place these three variables in a causal SEM model and test how well the model is confirmed.

We can depict complex relationships in the form of a diagram, as we did with path analysis (SEM can be viewed as an advanced version of path analysis), as well as algebraic equations. There are as many equations as there are endogenous (dependent) variables in the model. Endogenous variables are caused by exogenous (independent) variables. Exogenous variables are presumed to be caused by factors outside the model. We state a multiple regression model for each of the endogenous variables. Thus in the above example:

Endogenous variable = Function of (Exogenous variables)

Y = b1X1+ b2X2 + b3 X3 + e
RRSP buying = b1(income)+ b2(age) + b3(attitude to retirement) + error

If we have 10 different products and assume that different variables contribute to the buying of different products then we will have 10 such equations. In each equation the buying of each product is assumed to be the sum of all variables, each variable weighted according to its contributions to the buying behaviour. As before, the b coefficients are the regression coefficients and indicate the amount of change required in that exogenous variable to cause a unit change in the endogenous variable. 'e' is the error (or disturbance) term associated with unspecified and unspecifiable variables on the endogenous variable.

Manifest and latent variables
A major difference between path models and structural equation models is the extensive use of latent variables in SEMs. A variable is *manifest* if it is measured directly. For instance, annual income is a manifest variable in that it can be measured and verified by objective procedures. Some variables are *latent* . They are only measured indirectly. Brand loyalty is an example of a latent variable. It can only be inferred from observed variables. Depending on how it is defined, a person may or may not be brand loyal. Sometimes the same variable can be manifest or latent, depending on how it is measured. For instance, if a 'buyer' is defined as anyone who has bought the product in the past 12 months, 'buyer'

is manifest variable. If, on the other hand, a 'buyer' is defined on the basis of responses to a series of questions such as "How many times in the past year did you buy this product?", "Do you intend to buy this product in the next 30 days?" and "Would you buy this product even if cheaper alternatives are available?", then 'buyer' is a latent variable. Exhibit 1 and 2 illustrate the difference using an example.

Latent variables are subject to more measurement errors
We may also want to note that when we measure a variable directly (as when we ask whether a person has bought the product in the past 30 days), only one measurement error term is attached to the question. When we measure the same variable indirectly, the error terms increase. For instance, each one of the questions that help us define a 'buyer' is subject to measurement errors. So in our example there are six measurement error terms, one attached to each variable being measured.

SEMs are designed to handle latent variables.

Exhibit 1

Here, 'buyer' is a manifest variable. A buyer is defined as anyone who has bought the product in the past 30 days. This can be objectively and consistently measured.

Did you buy this product in the past 30 days?

YES ==> BUYER

NO ==> NON-BUYER

Exhibit 2

Here, 'buyer' is a latent variable. A buyer is defined as someone who answers positively to a certain number of questions. The questions and the extent of agreement required to qualify as a buyer are arbitrarily defined by the marketer.

- Did you buy this product in the past 30 days?
- Do you intend to buy the product in the next 30 days?
- Would you buy this product even if a similar product is available at a cost that is 10% lower?
- Have you been buying this product for, at least, the past 12 months?
- Did you switch to this product in the past 12 months?
- Do you intend to continue buying this product for the next 12 months?

Answered "YES" to at least 4 questions ==> BUYER

Answered "YES" to less than 4 questions ==> NON-BUYER

Statistical analysis

The type of variables used in SEM decides the statistical procedure used for analysis. For manifest variables, the Ordinary Least Squares (OLS) Regression technique is used since this technique requires rigorous assumptions. For latent variables, less stringent Maximum Likelihood Estimation (MLE) procedures are used. Most programs designed to carry out SEMs such as LISREL 7, EQS, and CALIS use Maximum Likelihood estimation procedures.

The main advantage in using the maximum likelihood procedures is that they, unlike OLS procedures, do not assume that variables are measured perfectly. Consequently, MLE account for measurement errors in the SEM. This minimizes the distorting effect of measurement of causal inferences found in OLS estimation techniques. In addition, the MLE procedures offer the following advantages:

- MLE does not assume recursiveness or one-way relationships. This means that we can hypothesize feedback and thus test the non-recursiveness of relationships among variables.
- MLE allows for the possibility that outside factors could influence the model. Consequently, it can model unspecified causes on endogenous variables by specifying an error or disturbance term associated with endogenous variables.

MLE makes some assumptions about the nature of the data. The first one is that the errors in endogenous and exogenous variables are uncorrelated. The second assumption is that the correlation matrix of the dependent variables is non-singular.

In the next article, we will discuss how the model is set up and analyzed in practice.

Book Review

A Manager's Guide to Marketing Research

by Ronald H. Rotenberg
Published by Harcourt, Brace & Company, Canada 414 pp. $36.95+GST

While there are many books on marketing research, there are few that are written specifically for managers. The manager entrusted with the responsibility of buying research either has to wade through tomes written for practitioners or has to rely solely on the advice of research houses and consultants whom the manger is expected to evaluate! So it is a pleasure to see Professor Rotenberg's book on marketing research which is written with the manager in mind.

The book has a number of attractive features: It is clearly written and well laid out. It covers topics of interest to research buyers: research process and design, how to choose and use research suppliers, basics of qualitative and quantitative research, questionnaire design, sampling, and how fieldwork is done. In addition, the book has a section on research ethics and report writing. Another useful feature of this book is 'The manager's checklists' (interspersed throughout the book) which lists items that a manager should take care of.

It is unfortunate that this book lacks an index. Not having an index can be an irritant to users. The user often needs to find information on a given topic and an index could make the job a lot easier.

There are statements in the book with which the present reviewer disagrees. For instance, Rotenberg defines sampling error as the "difference between the result obtained from a sample of respondents and the true population parameter". In reality, the difference between the obtained and 'actual' contains non-sampling errors as well. Sampling error will coincide with the difference only if no other errors are present, which is rarely the case.

There are statements in this book that puzzle the reviewer. The author suggests that January, February, July and August are slow months for research and that suppliers may lower their prices during these periods to keep people employed. While this is possible, I haven't found such differences in quotes. Prices seem to depend on the competitive environment rather than the volume of work that a company handles at any given time.

There are also statements in the book that delight the reviewer. Two examples are, Rotenberg's observations that "evaluating research suppliers be undertaken before a request for proposal is issued" and that it is unethical to ask for a RFP from a large number of suppliers. If these suggestions are followed, both the buyers and the providers of research will benefit.

A Manager's Guide to Marketing Research is a welcome addition to the field of marketing research literature for many reasons, the most important of which is that not many books seem to look at research from the manager's viewpoint.

Chuck Chakrapani

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