Market modelling · 4
Path Analytic Models - 1

Chuck Chakrapani

Observed relationships vs. causal connections
In research, we constantly identify relationships-consumption is related to age, income is related to education, social attitudes are related to voting behaviour and so on. An implied assumption is that relationships imply causation.

However, observed relationships do not guarantee causal connections. Consider the following two examples. If you observe that your product is consumed more by upper income earners you may expect consumption of your product to increase as consumers' income increases. This expectation is not unreasonable. Consider another situation however in which you observe that your product is consumed more by people who live in Vancouver . How reasonable is it to expect that the consumption of your product will grow as the population of Vancouver grows?

The above examples illustrate the difficulty we have in moving from observed relationship to causality. In a broad sense, we can never be sure of 'causality'. Causality is what we infer from observations and as such can never be proved by any number of supporting observations. Yet causality can be disproved by a single contradictory observation.

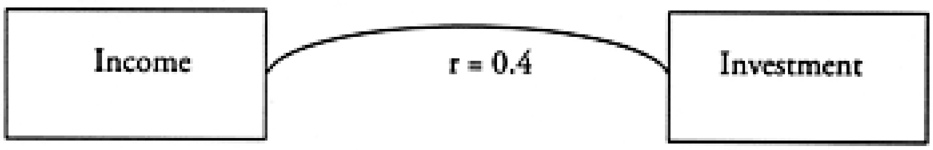
Causality as a working hypothesis
In spite of the above limitation, as marketing and research professionals, we have to assume causality, at least as a working hypothesis. For instance, we assume that promotional efforts 'cause' sales, otherwise we would not spend resources on promotional efforts. Even if all causal explanations are tentative and are no more than working hypotheses, there is still some value in such causal explanations.

Path analytic models force the researcher to make the implied causal assumptions explicit. There are many advantages to this approach. Some of these are:

- We can test the assumptions and the strength of the relationships through statistical techniques.
- We can test alternative models.
- We can identify variables that do not have direct impact. For instance, if we find a relationship between product consumption and residence in a particular region, path analytic models may identify the 'cause' of consumption to be one's income rather than one's residential status.

Path analysis
Let us start with the observed relationship between two variables in a survey. We note that a person's income and the value of his or her investments has a correlation of 0.4. We can represent this relationship as a curved line (see Exhibit 1). The curved line implies a simple relationship. It does not state the direction of the relationship or even whether the two variables are causally related at all.

Exhibit 1 Correlated Variables

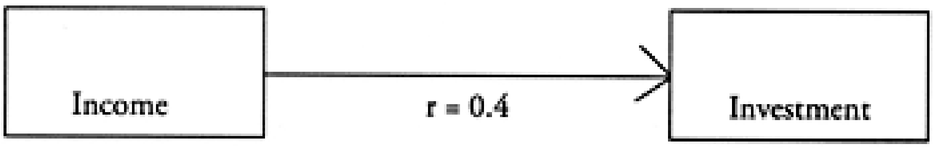


Weak causal ordering
In the next step, we begin to build the model by stating the direction of the relationship. For instance we may

state that higher income leads to (or 'causes') larger investments. Even here we don't state that higher income leads to a larger investment portfolio. Rather, we ask: if a higher income leads to higher level of investing, what is the nature of this relationship? This is called weak causal ordering. In weak casual ordering we take the responsibility for the model. The mathematics simply provides a way of formalizing our model.

We can state our model that higher income leads to (or 'causes') larger investments as a straightline arrowhead from the 'cause' to the 'effect' (see Exhibit 2).

Exhibit 2 Weak Causal Ordering



Exogenous and endogenous variables

While the model states that higher income leads to larger investments, it does not state what 'causes' income. The 'cause' of income comes from outside the model. Therefore income is called the exogenous variable.

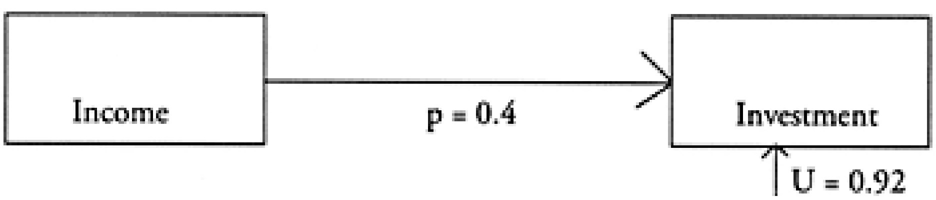
Investments, on the other hand, is 'explained' (at least in part) by income, and therefore its explanation is within the model. As a result, investment is called the endogenous variable, even though the explantion may only be partial.

The meaning of the diagram

The path diagram shows the correlation between a person's Income and Investments. A correlation of 0.4 means if a person's Income changes by one standard deviation, his or her Investments will change by 0.4 standard deviation. We can also state that Income 'explains' 16% of the variance in Investments. (16% is derived by squaring the correlation and multiplying the result by 100: 0.4 x 0.4 x 100 = 16%).

If Income explains 16% of the variance, it follows that 84% of the variance is still unexplained. To complete the model we introduce another path that shows the contribution of unexplained variables to Investments. The correlation between the unexplained variables and income is calculated by (unexplained variance), or (1-r2). In this instance, the unexplained variance works out to .92 ([1- 0.402]). This is shown in Exhibit 3.

Exhibit 3 Unexplained Variance



Extending the model

So far, we have not achieved anything new. We just represented the explained and unexplained relationship between two attributes and made our assumption of causality explicit. While this may be helpful, it does not add anything significant to our understanding. Path analytic models become more useful as we add more variables.

Suppose we find that Education shows a correlation of 0.5 with Investments. Now the relationships become more complex. It may be that higher education 'causes' a person to invest more; it may also be that higher education 'causes' higher income, which in turn 'causes' larger investments. To assess the relationship between education and income, we can compute the correlation. Let us say this correlation is 0.5.

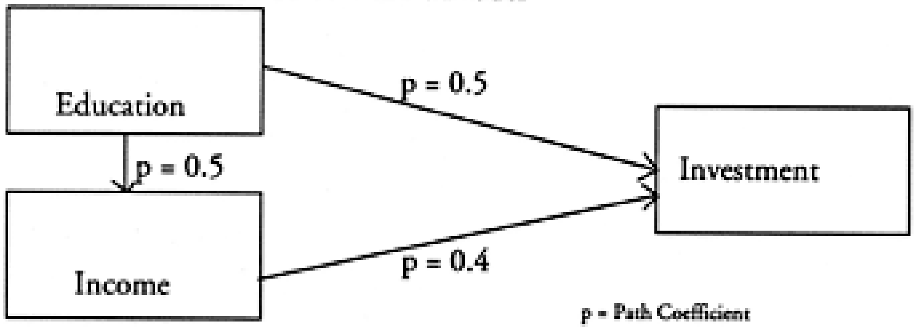
A recursive model

We can now build a model that states that:

- Higher education results in higher investments.
- Higher education results in higher income.
- Higher income results in higher investments .

This model is illustrated in Exhibit 4 and is called a recursive model. (It is a complete recursive model since no path is omitted in the diagram.)

Exhibit 3 3-Variable Recursive Model



Direct and indirect relationships

In the above example, the relationship between Education and Income is straightforward because no other variable is involved. The other relationships are somewhat more complex. Although Income has a correlation of .4 with Investments, only a part of this is directly causal. The other part is influenced by Education. We cannot use simple correlations where the relationships are more complex. For the model to be meaningful, we need to identify how much of the influence exerted by Income is *directly causal* and how much of it is *indirectly causal* (came about because Education influences Income).

Since we cannot use correlations as path coefficients when the relations include both direct and indirect paths, we use regression analyses instead. The reason for this is that the coefficients provided by regression analysis are 'partial regression coefficients'. This means that if we regress Investment on Education and Income, the coefficient we get are for Education if Income is kept constant and, for Income, if Education is kept constant. Thus the partial coefficients tell us the absolute contribution of each variable.

When a path model contains both direct and indirect paths, path analysis uses the beta weights (rather than the correlations) as path coefficients. In our example, we would run two regressions:

Dependent Variable	Independent Variables
1. Income	Education
2. Investment	Education Income

We will not run a regression of Income on Education because Education is an endogenous variable in our model. Suppose that our regression analysis produced the coefficients listed in Exhibit 5.

Exhibit 5

Regression	Dependent	Independent	Betas
1.	Income	Education	0.5
2.	Investment	Education Income	0.4 0.2

This information enables us to understand the nature of the relationships better, as shown in Exhibit 6.

Exhibit 6

Attributes	r	Direct Effect 1	Indirect Effect 2	Total Effect 3	Non-Causal 4
Edn. & Inc.	0.5	0.5	-	0.5	0.0
Edn. & Inv	0.5	0.4	0.1	0.5	0.0
Inc. & Inv.	0.4	0.2	-	0.2	0.2

1 *Beta coefficients from the regression equation*

2 *Derived by multiplying path coefficients along the way*
eg: Edn & Inv. = (Edn. & Inc) x (Inc.& Inv.) = .5x.2=.1

3 *Direct plus indirect effects.*

4 *Relationship that is not explained directly or indirectly*

From the exhibit one can glean several interesting relationships:

- The initial correlation between Education and Investment was 0.5. This relationship can now be broken down into two parts: Direct relationship (0.4) and Indirect relationship (0.1). The indirect relationship comes about because education influences income which in turn influences Investments.
- The initial correlation between Income and Investments was 0.4. Of this, only 0.2 can be attributed to direct effects. There are no indirect effects for this variable in the model. This means that 0.2 of the correlation is non-causal or 'spurious'.

Partitioning the relationship
Even the simplified example above which uses just three variables demonstrates how path analysis can identify direct relationships, indirect relationships and non-causal relationships in the marketing context. Its value increases when we add more variables and attempt to understand the complex relationships among a variety of marketing variables.

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