Journal of Information Technology Management

ISSN #1042-1319

A Publication of the Association of Management

THE USE OF CAUSAL ANALYSIS TECHNIQUES IN INFORMATION SYSTEMS RESEARCH: A METHODOLOGICAL NOTE

AL BENTO
UNIVERSITY OF BALTIMORE
abento@ubalt.edu

REGINA BENTO
UNIVERSITY OF BALTIMORE
rbento@ubalt.edu

ABSTRACT

Causal analysis techniques have been widely used in business research, but many Information Systems (IS) researchers are still unfamiliar with this methodology. This work describes a technique for causal analysis and discusses its rationale, assumptions and importance. The use of the technique is illustrated with two detailed examples of its application in IS research.

Keywords: research, methods, causality, ordinal data, non-parametric statistics

INTRODUCTION

Causal analysis techniques have been used extensively in other business disciplines, but their use in Information Systems research is more recent [1], [2], [6], [8], [9], [10], [11], [12]. Many IS researchers are still unfamiliar with this methodology, and our objective is to help bridge this gap by providing an introduction to the use of causal analysis. Here we describe a causal analysis technique, discuss its rationale, assumptions and importance, and use two examples from IS research to illustrate its application.

The progress of any field of knowledge depends on the ability of scholars to develop and test theories that satisfy the human need to understand causality. As Campbell reminds us, "a theory is a collection of assertions, both verbal and symbolic, that identifies WHAT variables are important for what reasons, specifies HOW they are interrelated and WHY, and identifies the CONDITIONS under which they should be related or not related." [5]

In order to develop and test theories that meet this definition, we need to engage more in the kind of systematic research that goes beyond the identification of correlations between variables, and actually explores causality. Much of MIS research is fragmentary in nature: a few variables are selected, and the author tries to find out whether and how they are related to each other, typically using simple correlations or more sophisticated tools like factor analysis, MANOVA, etc. Although this type of research can result in valuable contributions, it does not necessarily advance and test theory. Theoretical development requires methodologies that allow us to infer causality, and Herbert Simon warns us about the fallacy of interpreting correlation as causality [13].

Causal path analysis, the technique we explore here, is one of the most powerful methodologies to explore causality. Used extensively in the social sciences, but still sparingly in IS, causal path analysis is non-recursive in nature; it allows us to go beyond simply testing whether variables A and B are related, to say that A causes B, and that B does not lead to A. Given that most data in IS research is only nominal or ordinal in nature, rather than interval, the use of causal path analysis is particularly appropriate for IS scholars. Causal path analysis allows us to relax some of the stringent requirements or assumptions about the nature of the data and types of statistical distribution that are necessary in other approaches to causality, such as multiple regression and simultaneous equations.

Causal path analysis has been well developed and described by Simon [13] and Blalock [3] [4], among others. Our goal here is simply to summarize the technique in a series of steps and to explain its rationale, in order to help IS researchers interested in applying causal path analysis in their own research. For more in-depth familiarization with its terms, (such as levels of measurement, spurious relationships, mathematics of partial correlations, etc.), readers are referred to the work of Simon and Blalock.

In the sections that follow, we provide an introduction to causal path analysis and illustrate it using data from two previous research studies by the first author, which exemplify some of the earliest attempts to use this methodology in IS research. The material presented here has been used with great success in training doctoral students, junior faculty and even senior colleagues unfamiliar with the use of this particular methodology.

CAUSAL ANALYSIS TECHNIQUES

Causal analysis reasoning starts with the notion that "correlation analysis cannot be directly used to establish causality, because of the fact that correlations merely measure covariation or the degree to which several variables vary together" [3]. Science, however, requires the exploration of cause and effect. As part of the development of the body of knowledge in the social sciences, and specifically in business research, series of tools and concepts were developed and tailored to allow researchers to relate correlations to causality, using causal analysis techniques.

A series of assumptions are necessary for using causal analysis techniques with the appropriate rigor. The first of these assumptions is that all variables are measured in an interval scale. This is rarely the case in the social sciences, however, where most of the variables are measured in either ordinal or nominal scales. Researchers have chosen one of three approaches to this problem:

- ignore the issue completely, and assume that the tool is strong enough to yield acceptable results even with the imperfect information provided by ordinal and nominal scales.
- adapt or modify the technique by either: a) obtaining interval-like variables (e.g. converting nominal variables to yes-no, 0-1) and computing interval correlations; or b) maintaining the level of measurement of the variables and computing non-parametric (generally ordinal) correlations.
- consider the issue with "a grain of salt": compute both interval and non-parametric correlations and compare the results. If the use of interval methods has not introduced major distortions in the correlation coefficients computed, when compared with the non-parametric methods, proceed with the analysis assuming that the specific data set satisfies the conditions for interval data.

The second assumption is that only simple causal relations exist. Furthermore, the relationship between two variables is unidirectional, either going from X1 to X2, or from X2 to X1, never both ways. The relationship among variables is also a hierarchy, so that if we use subscripts to differentiate among the variables, no variable with a higher-numbered subscript can affect another with a lower-numbered subscript, as shown in Figure 1. This type of causal system is referred as a <u>recursive</u> system.

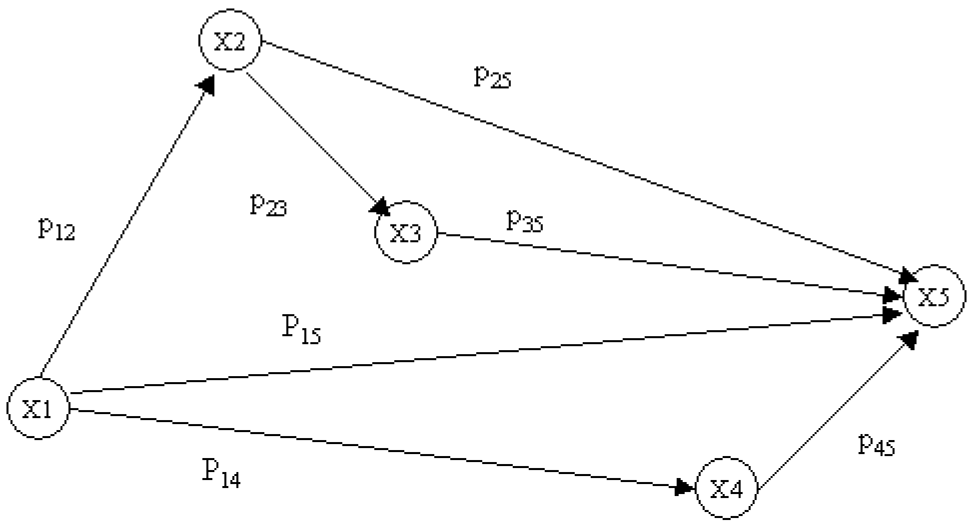

Figure 1: A causal recursive system

This assumption is more of a simplification of the hypotheses than a restriction. In computational terms, however, it has far-reaching implications. If we assume that the system is recursive, we can proceed to use Ordinary Least Squares (OLS) to estimate correlation coefficients, as in traditional regressions. On the other hand, if we assume that the relationship can be bi-directional and there is no hierarchy among the variables, then we have a case of a simultaneous-equation system requiring two- and three-stage estimation techniques, as discussed by Chin [7]. In practice, most researchers opt for the simplification provided by the causal analysis technique. It should also be noted that the use of OLS requires other assumptions not covered here [4].

Finally, causal analysis techniques make an important distinction between partial correlations and beta weights (or path coefficients). "The partial correlation is a measure of the **amount of variation explained** by one independent variable after the others have explained all they could. The beta weights, on the other hand, indicate **how much change** in the dependent variable is produced by a standardized change in one of the independent variables when the others are controlled" [4].

A typical causal path diagram, like the one shown in Figure 1, uses as path coefficients the beta weights, instead of partial correlation coefficients, to show the impact of one variable on the other(s). The variables are said to have direct and indirect effects on the other(s). For example, the **direct** effect of variable X2 on variable X5 is given by the beta weight or path coefficient P25. The **indirect** effect of variable X2 on variable X5 is obtained by multiplying the values of the path coefficients in the path between X2 and X5 through X3: P23 x P35. The **total** impact of X2 on X5 (or the correlation between X2 and X5) is obtained adding the direct and indirect effects: P25 + P23 x P35.

STEPS IN CAUSAL PATH ANALYSIS

The causal path analysis technique is a powerful and simple tool to explore cause and effect relationships among variables. One of its main advantages is that it uses as input the results of the readily available OLS and/or non-parametric correlation procedures contained in statistical packages such as SPSS, SAS, BMD, etc. Moreover, as discussed before, the assumptions it requires do not go much beyond those of Ordinary Least Squares.

Causal path analysis involves a set of very well defined steps:

Step 1. Define the Hypotheses as a Causal Recursive System

Variables should be organized following the rules described above for recursive causal systems. The order of the variables is established based on hypotheses inspired by a theoretical view-of-the-world, and justified in terms of previous research or of logical reasoning. Theoretical considerations should drive the choice of variables to be included or excluded, and how to display them in a graphical model. For an in-depth discussion of graphical modeling, please see Whetten on modeling as theorizing [14].

Step 2. Perform Data Collection

Data may be collected through a variety of procedures, such as interviews, surveys, field studies, secondary data, etc. The only constraints are the requirements imposed by OLS or non-parametric correlation procedure selected as suitable for the research.

Step 3. Compute Partial Correlation Coefficients and Beta Weights

The inputs for causal path analysis can be obtained from the standard results of statistical packages like SPSS, SAS, etc, as discussed previously. In practice, two approaches can be used in this step: non-parametric and parametric. In the non-parametric approach, the researcher assumes that the variables are nominal or ordinal, and computes only non-parametrical correlations (because beta weights depend upon the computation of standard deviation, an interval type of statistic). In the parametric approach, the researcher assumes that the variables are either truly interval, or that treating them as if they were interval has not introduced distortions to the computations of the correlation coefficients. Both partial correlation coefficients and beta weights are computed in the parametric approach.

Step 4. Draw a Causal Path Diagram with The Results

This is a modification of the original causal system created in Step 1 of the technique. Relationships found to be non-significant are dropped, and relationships that had not been hypothesized, but are logically justifiable and supported by the data, are added. If a non-parametric approach was used in Step 3, the partial correlations are shown in the diagram, and should not be interpreted as path coefficients. If a parametric approach was used in Step 3, the beta weights are shown in the diagram and interpreted as path coefficients -- i.e. effects on other variable(s).

Step 5. Compute and Represent in a Table the Direct and Indirect Effects

This step is only applicable in the parametric approach. The various paths in the causal system, from independent to dependent variables, are identified and the effects are computed.

In the next section we will illustrate the application of these steps to IS research, using both the non-parametric and the parametric approaches.

APPLICATIONS IN IS RESEARCH

The following examples are two of the earliest studies published in IS research using causal path analysis [1], [2]. The first, related to end-user computing (EUC), followed a non-parametric approach. The second, related to management of IS, used a parametric approach. For the sake of brevity, only a summary of the results of those two studies are presented here; please refer to the original studies for more details about their theoretical foundations, hypotheses, methodologies and implications.

Example of Non-Parametric Approach: EUC Study

The first study [1] proposed and tested a theoretical framework about the effect of End-user Computing Stage (1, Initiation; 2, Control; 3, Management), Strategic Impact of IT (1, Support; 2, Factory; 3, Turnaround; 4, Strategic) and Systems Development Level (1, Program; 2, Informal; 3, Formal; 4, Automated) on End-User Computer Policies (1, Limited; 2, Support; 3, Hygienic; 4, Comprehensive). Organization size (number of employees) and industry type (based on SIC code) were used as control variables.

The variables were all measured in ordinal scales or better, with the exception of one variable (Industry), which was measured in a nominal scale. Figure 2 depicts the variables used in the study, organized in a causal recursive system.

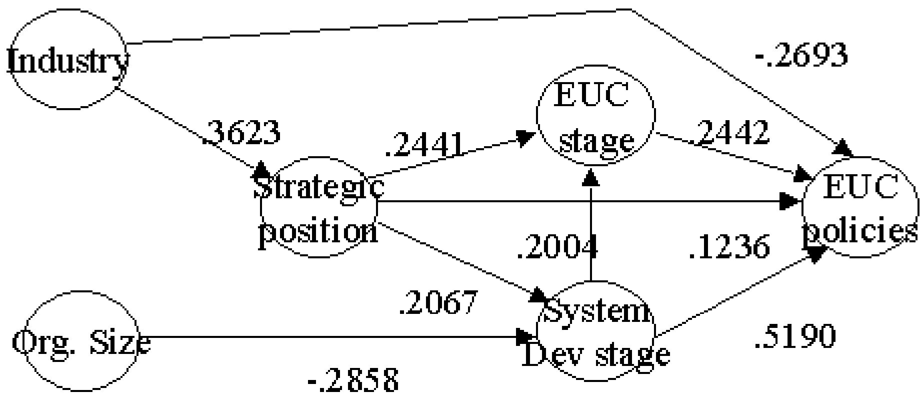

Figure 2: Causal Path Diagram with Ordinal Partial Correlations

The non-parametric correlation procedure of SPSS was used to compute the Kendall Tau-c simple correlation coefficients between the research variables. Partial correlations were computed to test the hypotheses. As shown by Blalock [3], the formulas for computing non-parametric partial correlations are the same as those used for computing the parametric partial correlation coefficients. Therefore, the New Regression procedure from SPSS was used for computing the non-parametric partial correlations by using as input the Tau-c correlation matrix previously obtained. The R^2 and F statistics computed by this procedure are meaningless in the case, as well as the significance level for partial correlation coefficients and beta weights [3].

The main results of this study are also shown in Figure 2. The numbers shown in the paths are partial correlation coefficients. The original simple correlations were all significant at .05 or better. Reading the path diagram in this case is more complicated. For example, the reasonably high partial correlation coefficient between Systems Development and End-User Computing Policies only indicates that when the level of System Development formalization increases, the level of EUC Policies formalization also does. This is neither the direct or the indirect effect of Systems Development on EUC Policies, but rather the amount of variation in EUC Policies explained by Systems Development, after all other variables have been used to explain the variation in EUC Policies.

Example of Parametric Approach: Study of Strategic IT Impact

The second study [2] proposed and tested a theoretical model to explain the type of impact that IT can have in organizations. The independent variables were IT Stage (1, Beginner; 2, Traditional; 3, Modern; 4, Advanced), CIO Management Style (1, Decisive; 2, Authoritative; 3, Participative; 4, Supportive), Management Model (1, Coordination; 2, Decision; 3, Action; 4, Analytical), Information Resource Management (1, Technical; 2, Mixed; 3, Managerial), Systems Development Level (1, Program; 2, Informal; 3, Formal; 4, Automated) and End-user Computing Stage (1, Initiation; 2, Control; 3, Management). The dependent variable was Strategic Impact of IT (1, Support; 2, Factory; 3, Turnaround; 4, Strategic). Control variables were organization size (total number of employees), industry type (based on SIC code) and MIS size (total number of employees in MIS).

The variables were all measured in ordinal scales or better, with the exception of one variable (Industry), which was measured again in a nominal scale. Figure 3 depicts the variables used in the research, organized in a causal recursive system, including only the paths that were later found to be significant.

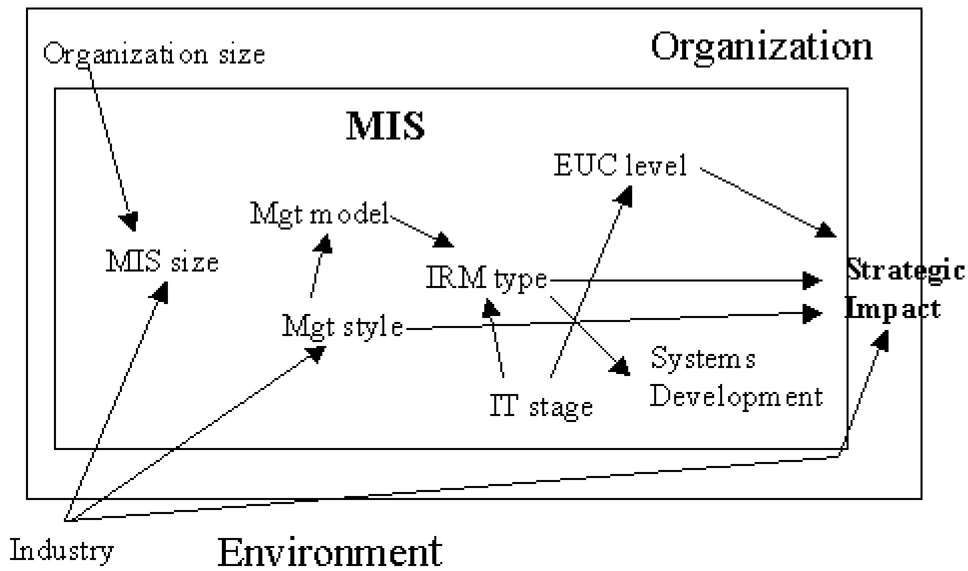

Figure 3: Causal recursive system model

The parametric and non-parametric correlation procedures of SPSS were used to compute the Pearson and Kendall Tau-c simple correlation coefficients between the research variables. The results were compared to check if the use of ordinal scales had introduced significant differences in the results between the two methods. The only cases where major differences occurred were in correlations between interval variables -- the Tau-c underestimated significantly the relation between the interval variables.

Finally, the SPSS stepwise multiple regression procedure was used to measure the overall strength of the relationships, and to compute beta weights, partial correlations and levels of significance. The level of significance chosen for the F statistic of the equations was .05. Therefore, new variables were included in the equation if, and only if, the F statistics of the equations were significant at least at .05.

Table 1: Causality Analysis of Strategic IT Impact

Variables	Partial correlation	Significance Level	Beta weights		
			Direct	Indirect	Total
Mgt. Style	.50572	.003	.72849	.18702	.91551
IRM type	.38861	.015	.46543		.46543
Industry	.39751	.014	.46913	-.09026	.37887
EUC level	-.26278	.081	-.29751		-.29751
Mgt. Model	...	n.s.	...	.28610	.28610

Adjusted R^2: .69644 Significance: .0008

Reading the results is much more straightforward in this case. Table 1 shows the path coefficients and an analysis of direct and indirect effects of beta weights. Management style, information resource management, industry and end-user computing level are the most important factors to explain directly and indirectly the strategic impact of IT (see Table 1). Management model was also found to have an indirect effect on strategic impact (see Table 1).

A series of other analyses for other variables could follow like Systems Development, End-User Computing, etc. The beta weights, or path coefficients, would indicate the effects of each of the other variables on the dependent variable considered.

FINAL COMMENTS

The aim of this work was to provide a description of the causal path analysis technique and illustrate its application in IS research. Causal analysis is a powerful and simple tool for researchers to explore cause and effect relationships and thus contribute to theory building in the IS field.

Causal path analysis is appropriate not only for interval data, but also for the nominal and ordinal data typically collected in IS studies (nominal, ordinal). Moreover, causal path analysis can be performed using the results of common statistical packages, such as SPSS, SAS, BMD, etc.

Causal path analysis can significantly help IS researchers move beyond the study of isolated variables and their relationships, and move on to the exploration of broader causal systems that can improve the quality of explanations in IS practice and theory.

REFERENCES

[1] Bento, A.M. “Policy Making for End-User Systems: A Model and Empirical Evidence,” *Journal of Information Technology Management*, Volume1, Number 2, 1990, pp 15-22.

[2] Bento, A.M & Grey, G. “Strategic Impact: an Empirical Study,” *Journal of Information Technology Management*, Volume4, Number 2, 1993, pp. 7-13.

[3] Blalock, Jr., H.M. *Causal Inferences in Non-Experimental Research*, University of North Carolina Press, Chapel Hill, 1964.

[4] Blalock, Jr., H.M. *Social Statistics*, 2nd.ed., McGraw-Hill, New York, 1979, pp. 462-477.

[5] Campbell, J.P. "The role of theory in industrial and organizational psychology." In Dunette, M.D. an Hough, L.M. (Eds)., *Handbook of Industrial and Organizational Psychology*. Consulting Psychologists Press, Palo Alto, CA, 1990, pp. 39-73.

[6] Chidambaram, L. “Relational Development in Computer Supported Groups,” *MIS Quarterly*, Volume 20, Number 2, 1996, pp 143-165.

[7] Chin, W.W. “Issues and Opinion on Structural Equation Modeling,” *MIS Quarterly*, Volume 22, Number 1, 1998, pp.vii-xvi.

[8] Ghani, J.A. & Chacko,C. “End-User Dependence on Computer Technology,” *Journal of Information Technology Management*, Volume 2, Number 3, 1991, pp. 9-15.

[9] Grant, R.A. “Building and Testing a Causal Model of an Information Technology's Impact,” *Journal of Information Technology Management*, Volume 2, Number 1, 1991, pp. 11-23.

[10] Nelson, K.M. & Cooprider, J.G. “The Contribution of Shared Knowledge to IS Group Performance,” *MIS Quarterly*, Volume 20, Number 4, 1996, pp.409-432.

[11] Nidumolu, S.R. & Grover, V. “The effect of Customizability and Reusability on Perceived Process and Competitive Performance of Software Firms,” *MIS Quarterly*, Volume 22, Number 2, 1998, pp.105-137.

[12] Pinsonneault, A. & Rivard, S. “Information Technology and the Nature of Managerial Work: From the Productivity Paradox to the Icarus Paradox?” *MIS Quarterly*, Volume 22, Number 3, 1998, pp. 287-311.

[13] Simon, H.A. (1954) “Spurious correlation: A Causal Interpretation,” *Journal of the American Statistical Association*, Volume 49, Number 264, 1954, pp. 467-479.

[14] Whetten, D. A. "Modeling-as-Theorizing: A systematic methodology for theory development." In David Partington (ed.), *Essential Skills for Management Research*. Sage, London, 2002, pp. 45-71.

AUTHORS BIOGRAPHIES

Al Bento is Professor of Information Systems at the Merrick School of Business, University of Baltimore. He is Editor-in-Chief of *the Journal of Information Technology Management* and Chair of the Association for Information Systems (AIS) Special Interest Group on Information Systems Security (SIGSEC). A graduate from UCLA, his current research includes projects on information security and performance measurement systems.

Regina Bento is Professor of Management and holder of the Hatfield-Merrick Distinguished Professor chair at the Merrick School of Business, University of Baltimore. She is past Chair of the Management Education and Development Division at the Academy of Management, and "OB-1" at the Organizational Behavior Teaching Society. A graduate from MIT, her current research includes projects on the behavioral aspects of the use of information technology.