Improving Our Understanding of the Conduct and Performance of Cooperative Businesses Using Directed Acyclic Graphs

Padmaja Pancharatnam, John Park, Amy Hagarman and Matt Murch

Abstract

A great deal of effort is devoted to the exploration of agribusiness firm behavior through survey instruments. Mainly in an attempt to understand the needs and desires of firms, such studies can improve the information available to marketers and other decision makers within an industry. The implications drawn from these studies might elicit discussions on "best practices" employed by successful firms. The data involved are typically a mixture of qualitative and quantitative variables and, not surprisingly, empirical analyses can give mixed results that are open to interpretation. A better understanding of the relationship among variables might improve structural analysis, however business theory is often vague in that regard. To help, we suggest the use of a less known methodology in the analysis of survey data, namely, the directed acyclic graph (DAG). We demonstrate its potential with an analysis of agricultural cooperative firms.

Introduction

Cooperative behavior might be understood utilizing the directed acyclical graph approach. The directed acyclical graph(DAG) examines the nature of causality between variables. Two applications of this approach to cooperative behavior are discussed. In the first instance DAG's are used to understand the factors that influence a cooperatives financial performance. In the second instance we would like to understand the relationship between the different variables that are said to influence a consumers willingness to adopt social media.

Most of the statistical techniques employed to test the strength of relationship between variables measure the degree of association rather than specifying the causal structures between variables. Often, the direction of causality between variables is of interest. The directed acyclical graph pictorially represents the causal structure between variables. The term directed implies that the direction of causation between two or more variables can be identified and the term "acyclic" implies that there is no path in which a variable can cause itself. The acyclic nature of the graph implies that the causal structure generated is robust. Certain criterion needs to be satisfied in order to generate a DAG. This condition is known as D separation. D separation implies that there exists no backdoor path between variables. One of the advantages of the DAG is that it helps choose the appropriate regression structure, and also provides an indication of whether the assumed regression structure is accurate. This helps in addressing the problem of endogenity among the variables.

Often, the relationship amongst multiple variables themselves might be of interest and with the aid of factor analysis many variables can be used to create indices and the relationship between these indices might be observed utilizing the directed acylical graph technique.

The Directed Acyclical Graph (DAG) technique is based on the following assumptions

For generating causal structure we assume that the following conditions hold: causal sufficiency, Markov parents' condition and faithfulness (Hagarman et al.2006)

Casual Sufficiency

The assumption of causal sufficiency implies that all variables that could cause the included variables have been specified. There are no omitted variables that could cause two or more of the included variables. In other words, there are no confounding effects or latent variables that could affect the included variables. This assumption can now be relaxed if certain conditions hold.

Markovian parents

The Markovian parent's condition implies that the probability of any variable can be expressed as a conditional distribution of its predecessors. This means that the probability of a variable in the graph may entirely be specified by its predecessors.

Faithfulness condition-

This condition states that if there is an edge between any two variables in the directed graph it implies that the variables are correlated. They do not have zero correlation. In contrast, if there is zero correlation between the variables there should be no edge between them. In a regression analysis it is assumed that the direction of causation between variables is known and the regression procedure reveals the strength and sign of the association between the explanatory variable and the dependent variable. Ganger causality is a method that represents the direction of causality between variables. Consider any two variables X and Y: there are five possibilities that can be represented utilizing directed graphs (Wang and Besssler,2008).

(1) X causes Y (→)

(2) Y causes X (←)

(3) They both simultaneously cause each other (↔)

(4)The causal flow cannot be determined by the information given in the sample (——)

(5)There is no causal relationship when the edges are removed

The Peter and Clarke (PC) algorithm is utilized to generate these diagrams. The first step is the generation of an undirected graph that connects each variable with all other variables that are present in the graph. The second step is to remove edges between correlated variables utilizing tests of correlation. The graph at this stage would still consist of undirected edges. In the third and final step markers are placed on one or both ends of the edges to indicate the direction of causality between the variables.

Directed graphs might also be employed to build regression model. The regression model is robust if the variables are D-separated. D separation exists between two variables X and Y if there is no other path through another variable Z. When the above condition is met it implies that the back door condition is satisfied. If such a back door condition does exist, it must be blocked by conditioning on the middle node or variable Z. By blocking back door paths the modeler prevents correlation between the explanatory or root variable X and the error term. Essentially, when a back door is left open, the effect of the variable is included in the error term. This implies that the explanatory variable X and the error term are correlated, since the variable Z causes both X and Y. So when some variable Z is the back door path variable the regression should be specified by the following equation (Hagarman et al. 2006)

(1) $Y = a + B_1X + B_2\ Z + e$

An application of the above concept is discussed in the following paragraphs. This application is interested in finding out how the various variables that measure the financial success of an agricultural cooperative are determined

The goal of this study is to evaluate the role of management philosophy in financial success of a firm. It is difficult to quantify and assess what actually is management philosophy. However, management philosophy influences various measures such as a manager's response to competition, government programs, financial condition of members, and other current events.

This application of the above theory consists of three steps.

1) Practices and perceptions that have a causal impact on the performance of cooperative are identified on the basis of a literature review.
2) Directed cyclical graphs are used to identify if there are any causal relationships between the management philosophy and performance and which perceptions and practices have the strongest influence.
3) The causal paths as identified by the DAGS are used to form regression models of performance drivers.

In order to generate a DAG certain casual assumptions have to be made about the variables included. Glymour, Scheines and Spirtes (1988) identified three primary ways in which a case could be made for the causal assumptions in a model. The first is a prior knowledge or a well justified theory that would imply a unique set of causal assumptions or reduce the number of alternative assumptions to a small number. This however, must be based on sound, severely tested theory. The second model is the use of experimental controls to isolate causal effects. As the authors point out, this is rarely feasible, particularly in areas such as economics. The final option is to use prior knowledge to conduct a systematic search for alternative models that is likely to provide the best available explanation of the data (Hagarman et al.2006).

The analysis conducted will utilize the first method; the variables selected for study are based on previous literature. Though the acyclical graph technique shows the direction of relationships between variables it does not reveal if the relationship between the variables is positive or negative. A regression analysis is utilized to determine the sign of the relationship between variables.

Data

The data for this analysis is based on a mail survey carried out in 2004. The survey contained information on financial and other general attributes. The survey was sent to 230 cooperatives in Texas, out of which 47 responded. This is a response rate of twenty percent which is considered to be acceptable. Information is collected on a number of variables in the following areas competition, strategic planning, equity and equity redemption, and member loyalty. Initially, a minimal DAG with just the financial and operations variables was generated. To this graph variables related to each of areas of interest was added separately. This resulted in a DAG for each of the areas competition, strategic planning, equity and equity redemption and member loyalty (Hagarman et al. 2006).

The objective of this study is to identify those perceptions and practices that separate successful, growing cooperative agribusinesses from stagnant ones. These measures are included to indicate that there are significant differences between the top performers and the bottom performers and it is possible to differentiate between them based on certain indicators. A graphical representation of these indicators has been presented below.

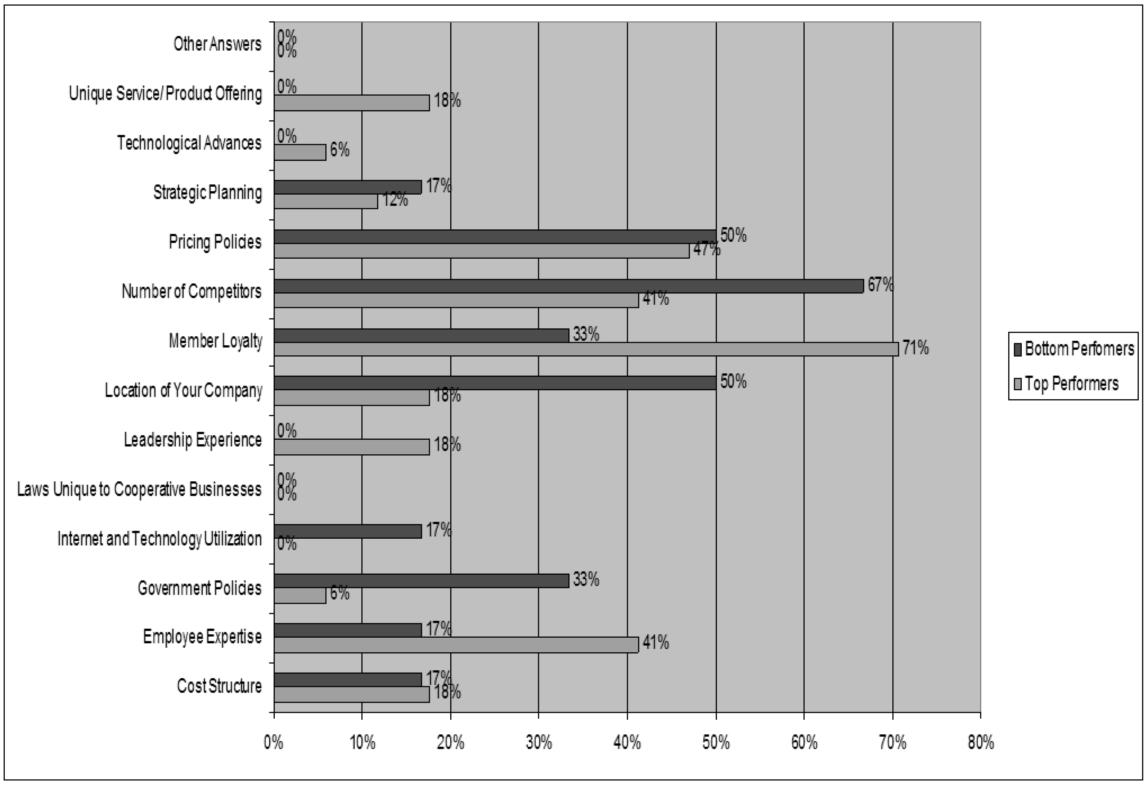


Figure 1 (Hagarman et al. 2006)

Results

Competition

Competition is hypothesized to have a negative impact on cooperative performance. There were qualitative measures as well as quantitative measures of competition. Examples of these measures are, the number of national chains that a company competes with (quantitative) and manager's perception of how government policies impact the cooperatives ability to compete (qualitative) (Hagarman et al 2005).

The DAG reveals that the basic firm model is significantly impacted by competitive variables. The causal sinks in the model are sales and average annual capital expenditures, whereas the number of business activities is a link in the causal chain, which ends in competition perception variables.

The regression analysis revealed that that the slope and intercept shifters were significant. We fail to reject the hypothesis that competition has a negative impact on cooperative performance.

Strategic Planning

Cooperatives that actively participate in strategic planning are said to experience greater financial success. The immediate benefits of strategic planning might not be seen in the short run, therefore even though a DAG is generated for this study, conclusions on this issue are deemed to be beyond the scope of this study.

Member Loyalty

Cooperatives that value customer loyalty are expected to be more financially successful. Data indicated that top performers are more concerned about member loyalty than others. While examining the member loyalty DAG, however it is apparent that membership and operational structures influence member loyalty to a greater extent. Further, while top performers intercept and slope shifters significantly explain the loyalty regression, the DAG reveals that the causal relationship hypothesized might be reversed (Hagarman et al. 2005)

Equity and equity planning

Cooperatives with greater understanding of equity planning are said to have a greater chance of financial success. There is a significant difference between the number of firms that formal equity redemption policies between the top and the bottom performers, however there appears to be little difference in perceptions of the management. The equity DAG was not too different from the Minimal DAG with financial and operations variables. In the regression analysis it was found that that the intercept and slope shifters were not significantly different from zero. Therefore, understanding of the equity position is rejected as a driver of greater financial performance (Hagarman et al 2005).

Discussion

The objective of the study is to identify the perceptions and practices that separate successful cooperatives from the stationary stagnant cooperatives. The statistical summary showed that there are differences between the top 10% of the cooperatives and the bottom ten percent of the cooperatives.

The differences between the top and bottom occur mainly in three areas.

1) Financial and operational form of business
2) Competition
3) Loyalty of members

The directed graphs also reveal causal relationships between performance factors and the impact of management perceptions on business. The competition DAG reveals that there are some incongruities when new variables are added. Government policies and cooperative leadership have the greatest impact on the competitive situation of a firm. The equity DAG is not too different from the basic DAG, however, being a regional cooperative appears to create changes and connections in the model. Strategic planning DAG does not differ too much from the original DAG. The Strategic planning variables intervene and create new sinks. This might imply that the physical size of the plant makes strategic planning more necessary. The final DAG generated relates to the member loyalty to the cooperative. Number of business activities the cooperative is involved in becomes the root cause in the con text of loyalty. Another interesting result from The DAG is that the amount of time a manager spends on his/ her greatest challenge is the root cause for every sink in the diagram.

The final conclusion is that human factors are at the beginning of the causal chain for every sink variable. This implies that the managers' perception is that loyalty begins with the people in the cooperative rather than with the price (Hagarman et al 2005).

The DAG’S provided a basis for regression analysis. With the aid of DAG’S it was possible to identify roots and sinks, and these provided the structural equations for simple regressions. The only interesting results were associated with models in which the dependent (sink) variables were sales. Intercept and slope shifters were jointly significant for each of the loyalty and competition models (Hagarman et al 2005).

Second Application

Often in a survey, in order to obtain information about particular attributes a large number of questions are asked. To understand the relationship between attributes, the dimensions of the data need to be reduced to create few meaningful indices. Factor analysis is a method used to reduce data and create these indices.

Factor analysis

Factor analysis is a method utilized to reveal the underlying constructs that affect a set of variables. The central idea on which it is based is that internal attributes affect surface attributes in a systematic fashion. These internal attributes are in some sense more fundamental than surface attributes. It involves a set of techniques designed to identify order and structure in data by providing parsimonious and meaningful explanations for the observed variation and covariation in surface attributes in a systematic fashion (Tucker and Macclum,1998). It can also be seen as a method that accounts for covariation amongst surface attributes. It is made up of two parts component analysis and factor analysis.

Practically, factor analysis is utilized to reduce the number of variables and it also used to create indices. For example in our study there are a number of questions about the various aspects of a cooperative, many of these responses might be correlated. Factor analysis leads to the creation of indices which are composed of correlated variables. It can be used to create a set of uncorrelated indices; these indices could then be used for analysis, thereby overcoming the problem of multicollinearity. However, if none of the original variables are correlated there is no reason to carry out factor analysis and there is nothing to explain. In this specific instance factor analysis is used to condense data on a number of questions on willingness of a firm to adopt social media. The factors that are created might be classified into the following categories.

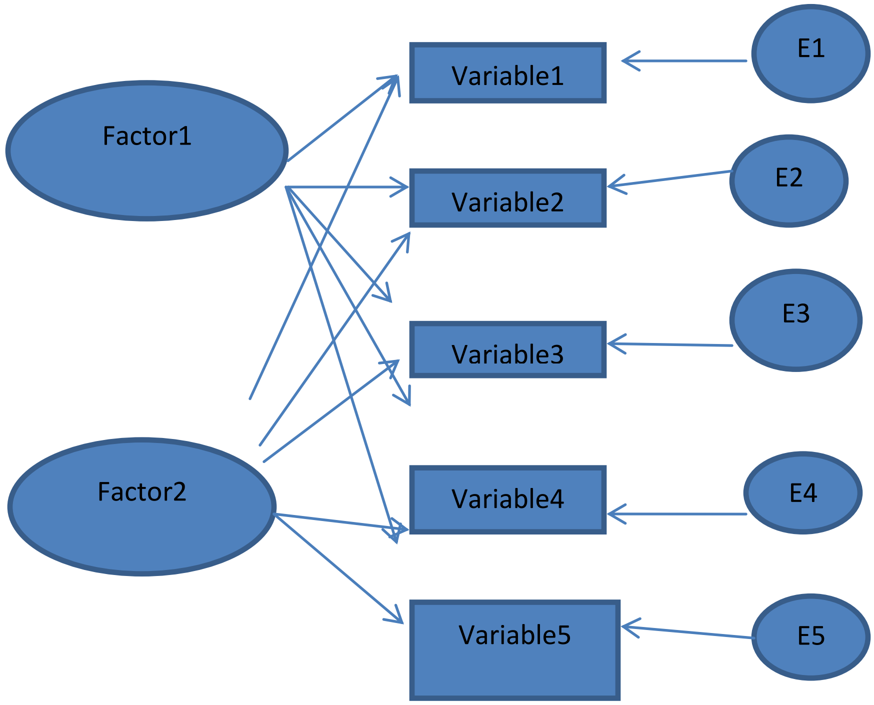
There are two types of factor analysis: confirmatory factor analysis and exploratory factor analysis. Confirmatory factor analysis is utilized to verify our initial assumptions about the factors that affect a set of variables. It is a test of whether a specified set of structures influence responses in a predicted way. Exploratory factor analysis on the other hand attempts to discover the nature of constructs that influence a set of responses. In this study we will be utilizing the exploratory factor analysis method to discover the variables that influence various responses on the state of an agricultural cooperative and their willingness to adopt technology. Exploratory factor analysis basically seeks to answer the following questions (Decoster, 1998).

1. How many factors are required to explain the pattern of relationships amongst these variables?

2. What is the nature of these factors?

3. How well do these factors explain the patterns in the observed variables?

4. How much of the purely random or unique variance does each variable include? The following diagram describes the relationship between variables in an exploratory factor analysis process



Source : DeCoster 1998

There are seven steps involved in carrying out factor analysis these are (DeCoster,1998)

1. Collect measurements- All measurements must be in the same units.
2. Obtain the correlation matrix- Obtain the correlation matrix between each of the variables
3. Select the number of factors for inclusion if there are K number of variables included in the model, there can be utmost K number of factors that account for these variables. A number of criteria might be used to extract the optimal number of factors. Amongst them is the Kaiser criterion, this is based on the number of Eigen values of the correlation matrix which are greater than one.
4. Extract the initial set of factors

   The correlation matrix must be submitted to a computer program to extract all the factors. This is extremely difficult to do by hand. There are many methods proposed for extracting the factors. Two of them are the maximum likelihood method and the principal factor method with iteration.

The maximum likelihood method is an iterative procedure and may therefore take greater time than the principal factor analysis method.

5. Rotate your factors for a final solution

   There are an infinite number of factors that account for the same amount of covariance in a given set of factors. However, some of the definitions are much easier to interpret than the others. Rotating the factors enables the research to provide the most meaningful interpretation of factors, while still accounting for the same amount of covariation amongst variables as the original specification. There are many types of rotation possible. The most commonly used are the orthogonal and the oblique rotation. The orthogonal rotation is used to create uncorrelated factors whereas the oblique rotation creates correlated factors.

6. Interpret your factor solution

   Each of the observed measures is linearly correlated to the factors. The strength of the relationship is expressed in the factor loading. The factor loading has the same interpretation as a regression coefficient, where the factor is regressed on the variables.

7. Construct scores for further analysis- If the facto are to be used for additional analysis, the corresponding factor scores need to created. The factor score is a linear combination of the variables that make up the factor weighted by the factor loading.

Factor analysis is often criticized because factor analysis solutions are often non unique. The process of factor analysis is not entirely an objective process; there is a great deal of subjectivity involved. It is up to the researcher to decide the number of underlying factors involved in explaining the correlation and covariation amongst variables, how these underlying factors are created and how they need to be interpreted, because of this subjectivity, critics of factors analysis feel that a researcher might be able to show that any factor of interest could be of relevance. However, practitioners of factor analysis disagree; they believe that there is some method and objectivity to the entire process.

Often factor analysis is confused with principal component analysis. They are however quite different. Principal component analysis is more concerned with the covariation between variables whereas factor analysis is interested in explaining the variability in the variables. In the principal component analysis method, the components extracted account for all the variance in the factors. The components are derived from the variables. Whereas, in factor analysis, hundred percent of the variance is not necessarily accounted for by the extracted factors. The components derived represent common latent factors and explain covaration amongst the variables. Exploratory factor analysis is utilized when the goal of the analysis is to reveal the structure of the underlying factors, principal component analysis is more of a data reduction technique (DeCoster 1998).

Once the factors are created the directed acyclical graph approach is utilized to understand the nature of nature of causality between two or more variables

The questions in this study were numerous. Many of these variables were measured on a likert scale of zero to one. There were also certain variables were continuous. These variables were classified into age and experience variables and attitude variables. Each of the factors created are discussed below. These variables were then utilized within the DAG framework to understand the nature of the causal relationship between variables. The factors that are created are the following and might be classified into the following categories. Factor analysis enables the creation of such indices. These indices are then utilized to understand the charecteristcs that motivate certain cooperative behavior. The relationship between these variables were also explored using DAGs. The management attitude variables appeared to be the roots. The logit regression also confirmed that these factors were the important in determing adoption of social media.

References

Hagarman, A., Park, J. and D. Leatham, Beyond Financials: How Management Philosophy affects cooperative performance (2006)

Bessler, DA Conference Notes 2012

DeCoster J., Overview of Factor Analysis http://www.stat-help.com/factor.pdf

Smith, W. notes for Stat 636 Statistics Department Tamu

Sprites, P., Glymour, C and R.Scheines (1993*). Causation, Predication and Search*, New York Springer-Verlag

Wang, Z., & Bessler, D. A. (2006). Price and quantity endogeneity in demand analysis: Evidence from directed acyclic graphs. *Agricultural Economics*, *34*(1), 87-95. doi:10.1111/j.1574-0862.2006.00106.x

Tucker,L. and R. MacCallum Exploratory factor analysis http://www.unc.edu/~rcm/book/factornew.htm

http://www.psych.cornell.edu/darlington/factor.htm