How Likely is Simpson’s Paradox in Path Models?

Ned Kock, Division of International Business and Technology Studies, Texas A&M International University, Laredo, TX, USA

ABSTRACT

Simpson’s paradox is a phenomenon arising from multivariate statistical analyses that often leads to paradoxical conclusions in the field of e-collaboration as well as many other fields where multivariate methods are employed. This work derives a general inequality for the occurrence of Simpson’s paradox in path models with or without latent variables. The inequality is then used to estimate the probability that Simpson’s paradox would occur at random in path models with two predictors and one criterion variable. This probability is found to be approximately 12.8 percent, slightly higher than 1 occurrence per 8 path models. This estimate suggests that Simpson’s paradox is likely to occur in empirical studies, in the field of e-collaboration and other fields, frequently enough to be a source of concern.

Keywords: E-Collaboration, Monte Carlo Simulation, Numeric Computation, Path Analysis, Simpson’s Paradox

1. INTRODUCTION

Simpson’s paradox, also known as the Yule–Simpson effect and the reversal paradox, is a phenomenon arising from multivariate statistical analyses (Pearl, 2009; Wagner, 1982). It is called a “paradox” because it often leads to paradoxical conclusions. Such conclusions may lead to the development of theories that incorporate causal effects disconnected from reality, based on empirical findings distorted by Simpson’s paradox. This applies to the field of e-collaboration as well as many other fields where multivariate methods are employed.

In the following sections, we provide an illustration of Simpson’s paradox in a path model, followed by a discussion of past estimates of the likelihood of Simpson’s paradox in contingency tables. We then provide a mathematical definition of Simpson’s paradox in path models, in the form of a basic inequality, which we use as a basis for the development of a more general Simpson’s paradox inequality that can be used in numeric estimations of the paradox’s probability. Finally, we illustrate the use of this general inequality to estimate the probability that Simpson’s paradox would occur at random in three-variable path models, which is found to be 12.8 percent; slightly higher than 1/8.
2. A PATH MODEL ILLUSTRATION OF SIMPSON’S PARADOX

Let us assume that we collected data from 300 firms about two variables: degree of collaborative management \((X)\) and firm success \((Z)\). The variable degree of collaborative management \((X)\) measures the degree to which managers and employees collaborate to continuously improve their firms’ productivity and the quality of their firms’ products. The variable firm success \((Z)\) measures the profitability of each firm.

Figure 1 shows a simple path model relating these two variables. Since this path model contains only two variables, then \(p_{XZ} = r_{XZ} = 0.5\); where \(p_{XZ}\) and \(r_{XZ}\) denote the path coefficient and the correlation between the two variables.

Figure 2 shows a slightly more complex path model with an additional variable pointing at \(Z\): degree of e-collaboration technology use \((Y)\). This new variable measures the degree to which an e-collaboration technology is used. The technology facilitates collaborative management is available in all firms studied. Because of this, firms where the degrees of collaborative management \((X)\) are high tend to also use the e-collaboration technology intensely, and thus present high degrees of e-collaboration technology use \((Y)\); hence the link \(X \rightarrow Y\) in the model.

In this example, the addition of the new variable led the path coefficient \(p_{XZ}\) for the link between the variables degree of collaborative management \((X)\) and firm success \((Z)\) to assume a negative value (-0.2), in contrast with the positive correlation \(r_{XZ}\) (0.5) between the same variables. This sign reversal characterizes what is known as Simpson’s paradox in path models.

Simpson’s paradox is generally perceived as a problematic phenomenon, since it leads to paradoxical conclusions based on empirical research (Pearl, 2009; Wagner, 1982). In the three-variable path model illustration above, the results would lead many researchers to believe that the association between degree of collaborative management \((X)\) and firm success \((Z)\) is negative.

The perception that Simpson’s paradox is a problematic phenomenon has motivated Pavlides & Perlman (2009), in a seminal article on Simpson’s paradox, to estimate the probability that Simpson’s paradox would occur at random. They focused on contingency tables, and made various assumptions, also estimating other conditional probabilities. The probability that Simpson’s paradox would occur at random in contingency tables was found to be fairly low, in the neighborhood of 1 to 2 percent.

Contingency tables summarize results of multivariate analyses involving categorical variables. Often the categorical variables assume two values each, leading to 2 x 2 tables. Multivariate analyses involving such categorical variables frequently take the form of ANCOVA analyses. Analyses summarized via 2 x 2 tables could be seen as special cases of path analyses where some of the variables are dichotomous.

The probability that Simpson’s paradox would occur at random obtained by Pavlides & Perlman (2009) could lead empirical researchers to reasonably conclude that Simpson’s paradox is an unlikely phenomenon, which should occur only very rarely in empirical research.

However, Pavlides & Perlman’s (2009) focus on multivariate analyses involving categorical variables may have led to a probability estimate that is significantly lower than that for path analyses, which typically include variables measured on ratio scales. This possibility provided the motivation for the current study.
Millennial Teamwork and Technical Proficiency’s Impact on Virtual Team Effectiveness: Implications for Business Educators and Leaders

The Influence of Information Technology on Organizational Behavior: Study of Identity Challenges in Virtual Teams
[www.igi-global.com/article/influence-information-technology-organizational-behavior/53189?camid=4v1a](www.igi-global.com/article/influence-information-technology-organizational-behavior/53189?camid=4v1a)