Chapter 34
How Likely is Simpson's Paradox in Path Models?

Ned Kock
Texas A&M International University, 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.

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.

DOI: 10.4018/978-1-5225-1918-8.ch034

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 $p_{ZX}=r_{ZX}=0.5$; where p_{ZX} and r_{ZX} 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.

Figure 1. Two-variable path model

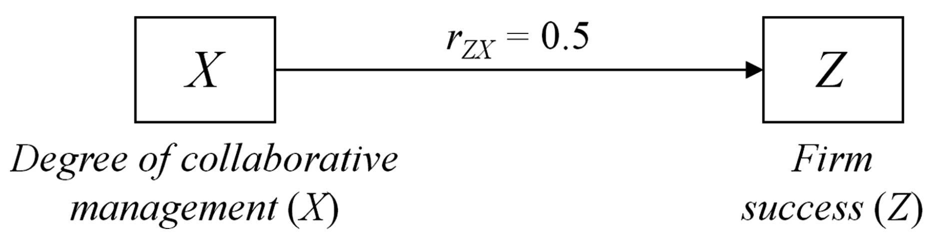
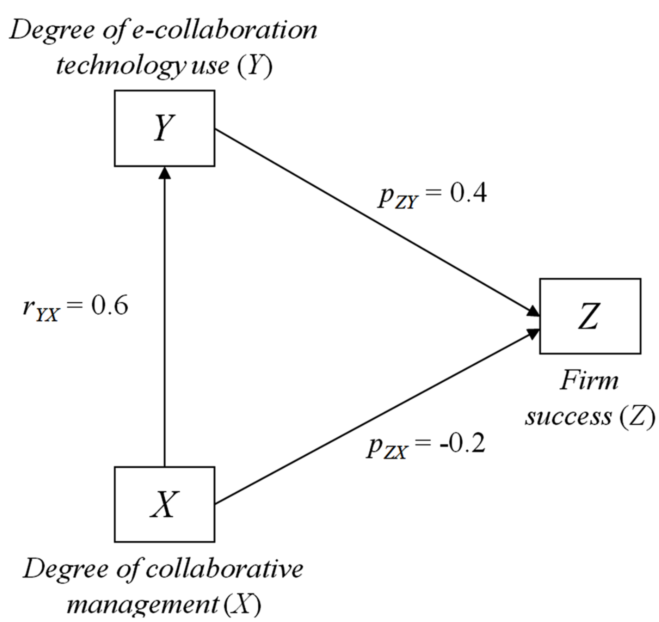


Figure 2. Three-variable path model



Related Content