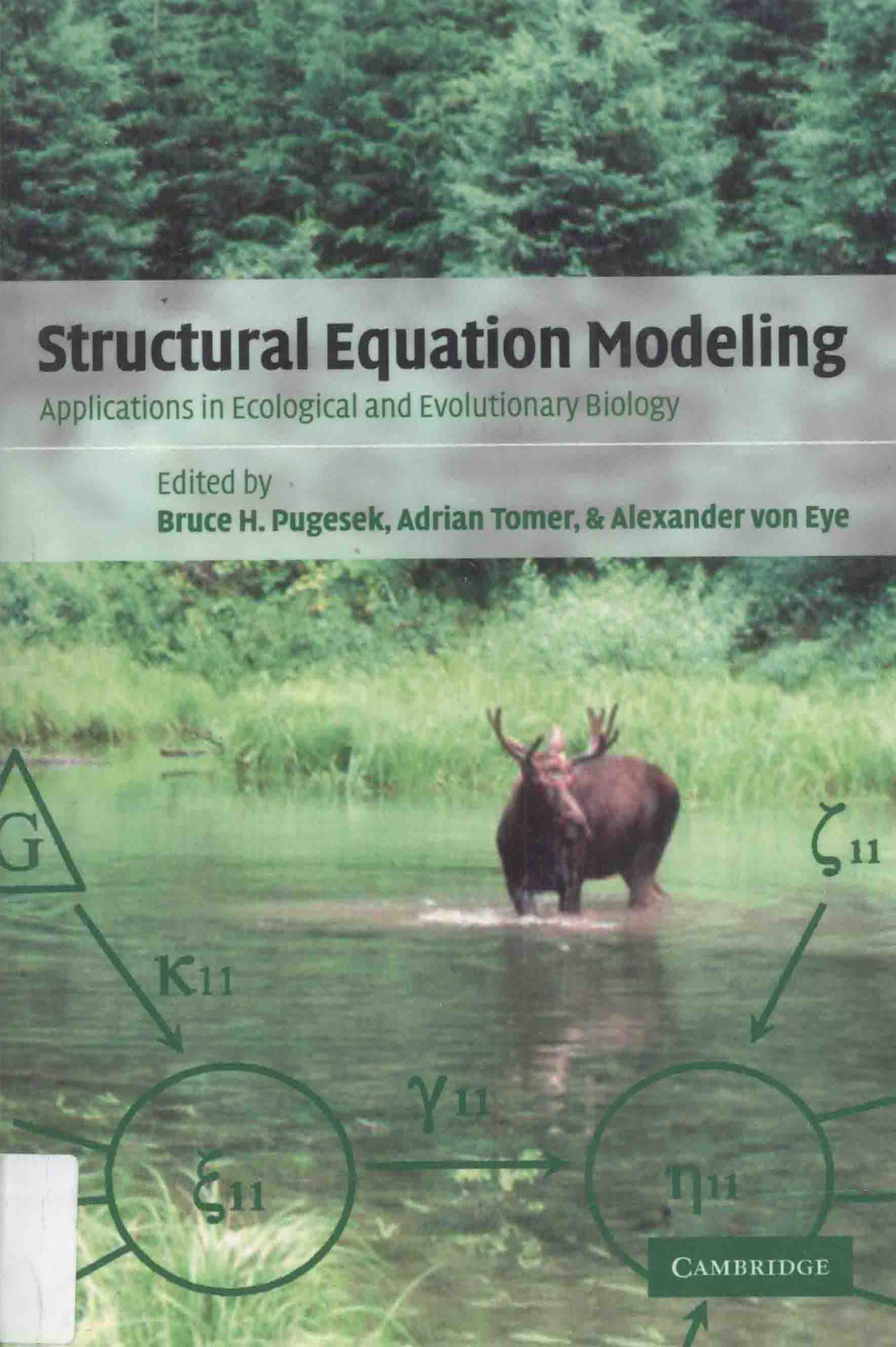
Structural Equation Modeling

Applications in Ecological and Evolutionary Biology

Edited by

Bruce H. Pugesek, Adrian Tomer, & Alexander von Eye



CAMBRIDGE

Structural Equation Modeling

Applications in ecological and evolutionary biology

Edited by

BRUCE H. PUGESEK

US Geological Survey – Biological Resources Division,
Northern Rocky Mountain Science Center, Bozeman

ADRIAN TOMER

Department of Psychology, Shippensburg University, Pennsylvania

and

ALEXANDER VON EYE

Department of Psychology, Michigan State University

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo, Delhi

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521104029

First published 2003
This digitally printed version 2009

A catalogue record for this publication is available from the British Library

ISBN 978-0-521-78133-6 hardback
ISBN 978-0-521-10402-9 paperback

Structural Equation Modeling
Applications in ecological and evolutionary biology

Structural equation modeling (SEM) is a technique that is used to estimate, analyze, and test models that specify relationships among variables. The ability to conduct such analyses is essential for many problems in ecology and evolutionary biology. This book begins by explaining the theory behind the statistical methodology, including chapters on conceptual issues, the implementation of an SEM study, and the history of the development of SEM. The second section provides examples of analyses on biological data including multi-group models, means models, P-technique and time-series. The final section of the book deals with computer applications and contrasts three popular SEM software packages. Aimed specifically at biological researchers and graduate students, this book will serve as a valuable resource for both learning and teaching the SEM methodology. Moreover, data sets and programs that are presented in the book can also be downloaded from a web site to assist the learning process.

BRUCE H. PUGESEK is a research statistician in the US Geological Survey – Biological Resources Division. He is the author of numerous scientific papers in the fields of ecology, behavior, evolution, and applied statistics.

ADRIAN TOMER is an Associate Professor at the Department of Psychology at Shippensburg University, Pennsylvania, where he teaches the psychology of aging and developmental psychology. He has a particular interest in the application of structural equation modeling to the behavioral and biological sciences.

ALEXANDER VON EYE is a Professor in the Department of Psychology at Michigan State University, where much of his research is dedicated to the development and application of statistical methods. He is the author of the book *An Introduction to Configural Frequency Analysis* (1990, ISBN 0 521 38090 1).

Contributors

Alexander von Eye
Department of Psychology
119 Snyder Hall
Michigan State University
East Lansing, MI 48824-1117
USA

Bret E. Fuller
Department of Public Health and Preventive Medicine
Oregon Health and Sciences University
3181 SW Sam Jackson Park Road
Mail Code CB-669
Portland, OR 97201-3098
USA

James B. Grace
US Geological Survey – Biological Resources Division
National Wetlands Research Center
700 Cajundome Blvd
Lafayette, LA 70506
USA

Patricia H. Hawley
Department of Psychology
Yale University
Box 208205
New Haven, CT 06520-8205
USA

Scott L. Hershberger
Department of Psychology
California State University – Long Beach
1250 Bellflower Blvd
Long Beach, CA 90840
USA

Peter S. Hovmand
Institute for Children, Youth, and Families
Kellogg Center, Suite 27
Michigan State University
East Lansing, MI 48824
USA

Bobby D. Keeland
US Geological Survey – Biological Resources Division
National Wetlands Research Center
700 Cajundome Blvd
Lafayette, LA 70506
USA

Todd D. Little
Department of Psychology
Yale University
Box 208205
New Haven, CT 06520-8205
USA

George A. Marcoulides
California State University, Fullerton
Department of Management Science
Langsdorf Hall no. 540
Fullerton, CA 92634
USA

Peter C. M. Molenaar
University of Amsterdam
Faculteit der Maatschappij – en Gedragswetenschappen, Afd. Psychologie
Roetersstraat 15
1018 WB
Amsterdam
The Netherlands

Makeba M. Parramore
Department of Human Development
Cornell University
Ithaca, NY 14853-2602
USA

Bruce H. Pugesek
U.S. Geological Survey – Biological Resources Division
Northern Rocky Mountain Science Center
1648 S. 7th Street, MSU
Bozeman, MT 59717-2780
USA

Michael J. Rovine
Director, Center for Development and Health Research Methodology
S-159 Henderson Human Development Bldg
Pennsylvania State University
University Park, PA 16802
USA

Bill Shipley
Département de Biologie
Université de Sherbrooke
Sherbrooke (Quebec)
Canada J1K 2R1

Adrian Tomer
Department of Psychology
Shippensburg University
Shippensburg, PA 17257
USA

Philip K. Wood
Department of Psychology
210 McAlester Hall
University of Missouri
Columbia, MO 62511
USA

Preface

This book describes a family of statistical methods known as structural equation modeling (SEM). SEM is used in a variety of techniques known as "covariance structure analysis", "latent variable modeling", "path modeling", "path modeling with LISREL", and sometimes it is mistaken for path analysis. This book will help biologists to understand the distinction between SEM and path analysis. The book consists of contributed chapters from biologists as well as leading methodologists in other research fields. We have organized the chapters and their content with the intent of providing a volume that readers may use to learn the methodology and apply it themselves to their research problems. We give the basic formulation of the method as well as technical details on data analysis, interpretation, and reporting. In addition, we provide numerous examples of research designs and applications that are germane to the research needs and interests of organismal biologists. We also provide, as a learning aide, the simulation programs, analysis programs, and data matrices, presented in the book at a website (http://www.usgs.gov/) so that readers may download and run them.

The book is divided into three sections. The first section, "Theory", describes the SEM model and practical matters of its application. Chapter 1 lays out the mathematics of SEM in a comprehensible fashion. Using an example from behavioral genetics, the authors express their model in what is called LISREL notation, a symbolic language that is commonly used to express SEM models. Chapter 2 describes SEM in a nonmathematical fashion. It will provide the reader with insight into how SEM differs from other methods and the benefits that may be obtained by using it. In Chapter 3, the author uses Huston's classic conceptual model of Shiras moose population dynamics to demonstrate that a complex model can be estimated and inferentially tested with SEM. The chapter provides examples of nonzero fixed parameters, measurement and structural models, and illustrates the distinction between exploratory and confirmatory models, the use of computer-generated information for model modification, and the concept and use of nested models. Chapter 4 provides a historical account of the development of SEM beginning with its origins in correlation and path analysis, and ending with the formulation of the LISREL model and its more recent expansion. Chapter 5 describes the numerous epistemological considerations that accompany an SEM study and provides guidelines for the implementation and reporting of SEM results. Details on the development of measurement instruments, sample size, model identification, fit indices, and other considerations necessary to the successful implementation of SEM are provided and well referenced. For those readers who wish only to understand SEM so that they can read and appreciate research that utilizes the method, we recommend that they read Chapters 1 through 3 of this section. For those who wish to implement an SEM study we recommend also Chapter 5. Everyone will benefit from reading the historical account of SEM, Chapter 4, especially those who seek a review of key papers past and present on the theoretical aspects of confirmatory factor analysis, maximum likelihood estimation, and other key components of the SEM methodology.

Section 2, "Applications", provides a sampling of the numerous ways that SEM can be employed. In Chapter 6, a confirmatory factor analysis of elephant behavior is presented. The authors provide examples of P-technique where data from a single individual are analyzed on a number of variables across a number of discrete points in time. Chapter 7 contrasts ordination techniques commonly employed in plant biology with an SEM approach. In Chapter 8, the author explores the notion of equivalent models in which more than one model may explain a data set. In Chapter 9, the author contrasts dynamic modeling, a method that is frequently employed in the study of complex ecological systems, with SEM. The strengths, limitations, and weaknesses of both methods are discussed. In Chapter 10, the author describes means modeling with SEM. Three examples of ANOVA applications, including, time-series analysis are presented. Chapter 11 addresses multigroup models, a method that allows comparisons of complex systems of variables from two or more groups. This approach has significant value for use in studies of multiple populations, habitat restoration, and situations where experimental versus control settings are desired at levels of organization such as the system or landscape level. Chapter 12 describes means modeling with latent variables. An example is provided for the study of natural selection in which environmental variables may impact phenotypic responses to a selection event. Chapter 13 provides an example of longitudinal analysis with SEM. The authors analyze tree growth data with

SEM and contrast results with an analysis of the data using latent growth curve methods.

Section 3, "Computing", contains Chapter 14, which discusses the relative merits of three popular software packages that perform SEM analysis. The authors compare performances on an analysis of R. A. Fisher's Iris data as well as compare features available in the software packages.

Contents

Section 1 Theory

1 Structural equation modeling: an introduction

Scott L. Hershberger, George A. Marcoulides,
and Makeba M. Parramore

Abstract

This chapter provides an introduction to structural equation modeling (SEM), a statistical technique that allows scientists and researchers to quantify and test scientific theories. As an example, a model from behavioral genetics is examined, in which genetic and environmental influences on a trait are determined. The many procedures and considerations involved in SEM are outlined and described, including defining and specifying a model diagrammatically and algebraically, determining the identification status of the model, estimating the model parameters, assessing the fit of the model to the data, and respecifying the model to achieve a better fit to the data. Since behavioral genetic models typically require family members of differing genetic relatedness, multisample SEM is introduced. All of the steps involved in evaluating the behavioral genetic model are accomplished with the assistance of LISREL, a popular software program used in SEM.

Introduction

Structural equation modeling (SEM) techniques are considered today to be a major component of applied multivariate statistical analyses and are used by biologists, economists, educational researchers, marketing researchers, medical researchers, and a variety of other social and behavioral scientists. Although the statistical theory that underlies the techniques appeared decades ago, a considerable number of years passed before SEM received the widespread attention it holds today. One reason for the recent attention is the availability of specialized SEM programs (e.g., AMOS, EQS, LISREL, Mplus, Mx, RAMONA, SEPATH). Another reason has been the publication of several introductory and advanced texts on SEM (e.g., Hayduk, 1987, 1996: Bollen, 1989; Byrne, 1989, 1994, 2000; Bollen & Long, 1993; Hoyle, 1995; Marcoulides & Schumacker, 1996; Schumacker & Lomax, 1996; Schumacker & Marcoulides, 1998; Raykov & Marcoulides, 2000), and a journal, devoted exclusively to SEM, entitled *Structural Equation Modeling: A Multidisciplinary Journal.*

In its broadest sense, SEM models represent translations of a series of hypothesized cause–effect relationships between variables into a composite hypothesis concerning patterns of statistical dependencies (Shipley, 2000). The relationships are described by parameters that indicate the magnitude of the effect (direct or indirect) that independent variables (either observed or latent) have on dependent variables (either observed or latent). By enabling the translation of hypothesized relationships into testable mathematical models, SEM offers researchers a comprehensive method for the quantification and testing of theoretical models. Once a theory has been proposed, it can then be tested against empirical data. The process of testing a proposed theoretical model is commonly referred to as the "confirmatory" aspect of SEM (Raykov & Marcoulides, 2000). Another aspect of SEM is the so-called "exploratory" mode. This aspect allows for theory development and often involves repeated applications of the same data in order to explore potential relationships between variables of interest (either observed or latent).

Latent variables are hypothetical or theoretical variables (constructs) that cannot be observed directly. Latent variables are of major importance to most disciplines but generally lack an explicit or precise way of measuring their existence or influence. For example, many behavioral and social scientists study the constructs of aggression and dominance. Because these constructs cannot be measured explicitly, they are are inferred through observing or measuring specific features that operationally define them (e.g., tests, scales, self-reports, inventories, or questionnaires). SEM can also be used to test the plausibility of hypothetical assertions about potential interrelationships between constructs and their observed measures or indicators. Latent variables are hypothesized to be responsible for the outcome of observed measures (e.g., aggression is the underlying factor influencing one's score on a questionnaire that attempts to assess offensive driving behavior). In other words, the score on the explicit questionnaire would be an indicator of the construct or latent variable – aggression. Researchers often use a number of indicators or observed variables to examine the influences of a theoretical factor or latent variable. It is generally recommended that researchers use *multiple indicators* (preferably more than two) for each latent variable considered in order to obtain a more complete and reliable "picture" than that provided by a single indicator (Raykov & Marcoulides, 2000). Because both observed and latent variables can be independent or dependent in a proposed model, a more detailed description of this issue will be provided later in this chapter.

Definition and specification of a structural equation model

The definition of a SEM model begins with a simple statement of the verbal theory that makes explicit the hypothesized relationships among a set of studied variables (Marcoulides, 1989). Typically, researchers communicate a SEM model by drawing a picture of it (Marcoulides & Hershberger, 1997). These pictures, or so-called *path diagrams*, are simple mathematical representations (but in graphical form) of the proposed theorical model. Figure 1.1 presents the most commonly used graphical notation for the representation of SEM models. As will become clear later, path diagrams not only aid in the conceptualization and communication of theoretical models, but also substantially contribute to the creation of the appropriate input file that is necessary to test and fit the model to collected data using particular software packages (Raykov & Marcoulides, 2000).

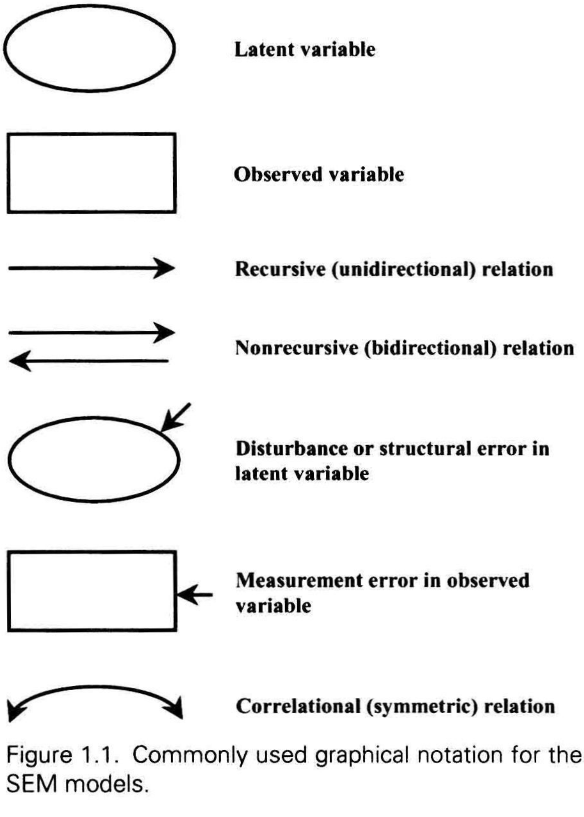


Figure 1.1. Commonly used graphical notation for the representation of SEM models.

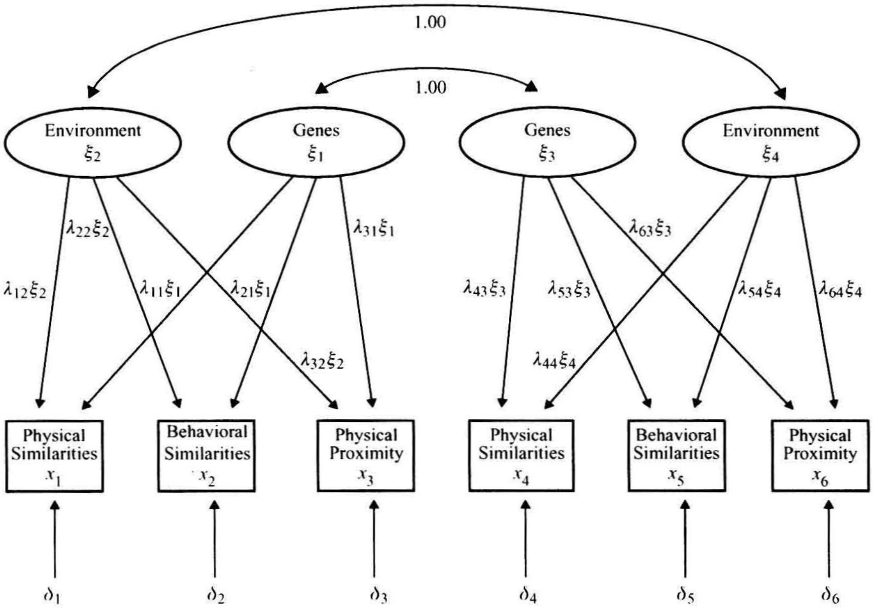


Figure 1.2. A model of sibling relatedness, in which the squares denote observed variables; the circles denote latent variables; the $\lambda_{ij}\xi_j$ are paths connecting latent with observed variables; and the δ_i are errors in the observed variables.

Figure 1.2 presents a simple example of a proposed theoretical model about sibling relatedness from the field of behavioral genetics. For years researchers have tried to understand the the "nature–nurture" phenomena by studying monozygotic twins, dizygotic twins, and nontwin siblings. To assess the amount of "relatedness" between siblings, researchers often use different types of questionnaire, standardized scales and tests, and independent observations. Two possible sources of relatedness between siblings are each sibling's genotype and environment. One may therefore define two different latent variables (i.e., genotype and environment) for each sibling, and denote each latent variable in the model by using the Greek letter ξ (ksi). Three possible observable variables (measures) of genotype and environment might be physical similarities, behavioral similarities, and physical proximity (Segal *et al.*, 1997). As it turns out, the scores or results observed for individuals on these variables will make up the correlation or covariance matrix that is analyzed to test a proposed model. The x values, which represent the observed variables or so-called *indicators*, are representative of the latent variables and make up the LAMBDA ($\mathbf{\Lambda}_x$) matrix. The error terms