

J. Christopher Westland

Structural Equation Models

From Paths to Networks

Studies in Systems, Decision and Control

Volume 22

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

About this Series

The series "Studies in Systems, Decision and Control" (SSDC) covers both new developments and advances, as well as the state of the art, in the various areas of broadly perceived systems, decision making and control- quickly, up to date and with a high quality. The intent is to cover the theory, applications, and perspectives on the state of the art and future developments relevant to systems, decision making, control, complex processes and related areas, as embedded in the fields of engineering, computer science, physics, economics, social and life sciences, as well as the paradigms and methodologies behind them. The series contains monographs, textbooks, lecture notes and edited volumes in systems, decision making and control spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

More information about this series at <http://www.springer.com/series/13304>

J. Christopher Westland

Structural Equation Models

From Paths to Networks



Springer

J. Christopher Westland
Department of Information and Decision Sciences
University of Illinois
Chicago, IL, USA

Library of Congress Control Number: 2015933613

Springer Cham Heidelberg New York Dordrecht London
© Springer International Publishing Switzerland 2015

© Springer International Publishing Switzerland 2015
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

This book evolved from a series of Ph.D. seminars presented to students both at Tsinghua University in Beijing and at the University of Illinois—Chicago. My lectures, in turn, grew out of my earlier investigation into sampling requirements in the sort of latent variable structural models that are prominent in social science fields like psychology, marketing, and information systems. In pursuit of my research, I discovered an academic domain steeped in mythologies and ambiguities, many that seem to have arisen during the fierce academic battles surrounding the Cowles Commission and Economics faculty at the University of Chicago in the 1940s. My home base in Chicago provided a unique opportunity to track down some of the original materials surrounding these debates. In the process, I gained an appreciation for the conflicting claims made by Cowles' systems of regression researchers, in contrast to those asserted by the isolated clique of Nordic researchers who single-handedly created the PLS path analysis and LISREL methodologies. The Cowles Commission ultimately prevailed in mainstream statistics and econometrics, but the Nordic approaches survived, and indeed thrived, in less rigorous arenas. This book surveys the full range of available structural equation modeling path analysis methodologies, from their early development in genetics to their current merging into network analysis tools. Applications of path analysis with structural equation models have steadily expanded over a broad range of disciplines—especially in the social sciences where many if not most key concepts are not directly observable, and the latent variable characteristics of these methods are especially desirable. This is the first book to extensively review the historical underpinnings that have defined the applications, methods, and assumptions invoked in each competing approach to estimation. History matters in understanding why particular academic disciplines have clustered around one or the other of these competing approaches. Knowing the background of PLS path analysis or LISREL is essential to understanding their strengths and weaknesses. The ability to accommodate unobservable theory constructs through latent variables has in particular grown with the expansion of the social sciences in universities. Latent variable constructs are fully explained here, and new methods are presented for extending their power. New techniques for path

analysis are surveyed along with guidelines for data preparation, sample size calculation, and the special treatment of Likert scale data. Tables of software, methodologies, and fit statistics provide a concise reference for any research program, helping assure that its conclusions are defensible and publishable. It is my hope that both scholars and students will find this book an accessible and essential companion to their research.

Chicago, IL, USA
February 2015

J. Christopher Westland

Contents

1	An Introduction to Structural Equation Models	1
1.1	The Problem Domains of Structural Equation Models.....	3
1.2	Motivation and Structure of This Book	4
	References.....	7
2	A Brief History of Structural Equation Models.....	9
2.1	Path Analysis in Genetics	9
2.2	Sewall Wright's Path Analysis.....	10
2.3	Networks and Cycles	13
2.4	What Is a Path Coefficient?.....	14
2.5	Applications and Evolution.....	15
2.6	The Chicago School.....	16
2.7	The Scandinavian School.....	18
2.8	Limited and Full Information Methods.....	20
2.9	A Look Towards the Future	20
	References.....	21
3	Partial Least Squares Path Analysis	23
3.1	PLS Path Analysis Software: Functions and Objectives.....	24
3.2	Path Regression.....	24
3.3	Herman Wold's Contributions to Path Analysis	26
3.4	Possible Choices for Path Coefficients: Covariance, Correlation, and Regression Coefficients	28
3.4.1	Covariance and Variance.....	28
3.4.2	Correlation	28
3.4.3	Regression Coefficients	28
3.5	Lohmöller's PCA-OLS Path Analysis Method.....	30
3.6	PLS Path Analysis vs. PLS Regression	32
3.7	Resampling	34
3.8	Measures	35

3.9	“Limited Information”	35
3.10	Sample Size in PLS-PA.....	36
	References.....	43
4	LISREL and Its Progeny	47
4.1	LISREL	47
4.2	LISREL Performance Statistics	53
	References.....	58
5	Systems of Regression Equations.....	61
5.1	The Birth of Structural Equation Modeling	61
5.2	Simultaneous Regression Equation Models.....	62
5.3	Estimation	64
5.4	A Comparative Example.....	67
5.5	A Model of Innovation, Structure, and Governance with Latent Variables in Systems of Regression Equations.....	75
	References.....	81
6	Data Collection, Control, and Sample Size.....	83
6.1	The Role of Data.....	83
6.2	The Ancient Roots of Model-Data Duality.....	84
6.3	Data: Model Fit.....	88
6.4	Latent Variables	92
6.5	Linear Models	94
6.6	Hypothesis Tests and Data	95
6.7	Data Adequacy in SEM	95
6.8	Minimum Sample Size for Structural Equation Model Estimation.....	99
6.9	Minimum Effect Size δ	102
6.10	Estimator for Correlation in a Bivariate Normal Distribution	104
6.11	Calculation of Sample Size on a Single Link	106
6.12	Can Resampling Recover Information Lost Through Likert Mapping?.....	109
6.13	Data Screening	110
6.14	Exploratory Specification Search.....	111
	References.....	112
7	Survey and Questionnaire Data.....	117
7.1	Rensis Likert’s Contribution to Social Research	118
7.2	Likert Scales.....	119
7.3	Fisher Information in Likert Items.....	125
7.4	Examples of Likert Scalings and Their Consequences	128
7.5	Design Trade-Offs in Likert Scaling.....	133
7.6	Known Unknowns: What Is a Latent Variable?.....	135
	References.....	136

8 Research Structure and Paradigms	139
8.1 The Quest for Truth.....	140
8.2 Research Questions	141
8.3 Models.....	141
8.4 Theory Building and Hypotheses.....	142
8.5 Hypothesis Testing.....	145
8.6 Model Specification and Confirmation	146
8.7 How Many Alternative Models Should You Test?.....	147
8.8 Distributional Assumptions.....	149
8.9 Statistical Distributions: Are They Part of the Model or Are They Part of the Data?	150
8.10 Causality	151
8.11 The Risks of Received Wisdom: A Case Study	152
8.12 Physics Envy and the Pretentious Model.....	153
8.13 Design of Empirical Studies	154
8.14 Significance Testing	156
8.15 Model Identification.....	156
8.16 Negative Error Variance Estimates	157
8.17 Heywood Cases.....	157
8.18 Empirical Confirmation of Theory.....	158
References.....	159
9 From Paths to Networks: The Evolving Science of Networks	161
9.1 Genetic Pathways Revisited.....	162
9.2 The Evolution of SEM Research Questions	164
9.3 The New Language of Social Networks	166
9.3.1 Visualization	166
9.3.2 Link Qualifiers and Metrics	167
9.3.3 Topological Statistics	167
9.3.4 Network Archetypes	168
9.4 Cross-Pollination: Biological Networks and Banking Ecosystems	169
9.5 The Future	170
References.....	171
Index.....	173

Chapter 1

An Introduction to Structural Equation Models

The past two decades have witnessed a remarkable acceleration of interest in structural equation modeling (SEM) methods in many areas of research. In the social sciences, researchers often distinguish SEM approaches from more powerful systems of regression equation approaches by the inclusion of unobservable constructs (called latent variables in the SEM vernacular), and by the use of computationally intensive iterative searches for coefficients that fit the data. The expansion of statistical analysis to encompass unmeasurable constructs using SEM, canonical correlation, Likert scale quantification, principal components, and factor analysis has vastly extended the scope and relevance of the social sciences over the past century. Subjects that were previously the realm of abstract argumentation have been transported into the mainstream of scientific research.

The products of SEM statistical analysis algorithms fall into three groups: (1) pairwise canonical correlations between pairs of prespecified latent variables computed from observable data (from the so-called partial least squares path analysis, or PLS-PA approaches); (2) multivariate canonical correlation matrices for prespecified networks of latent variables computed from observable data (from a group of computer-intensive search algorithms originating with Karl Jöreskog); and (3) systems of regression approaches that fit data to networks of observable variables. A fourth approach is fast emerging with the introduction of powerful new social network analysis tools. These allow both visualization and network-specific statistics that draw on an old and rich literature in graph theory and physical network effects.

Most of the PLS-PA algorithms are variations on an incompletely documented software package released in 1980 (Lohmöller, 1988, 1989) and some even use this 40-year-old Fortran code unmodified inside a customized user interface wrapper. To make matters worse PLS-PA is a misnomer—something its inventor Herman Wold tried unsuccessfully to correct—and is unrelated to Wold's (1973) partial least squares regression methods, instead being a variation on Wold's (1966, 1975) canonical correlation methods.

Two different covariance structure algorithms are widely used: (1) LISREL (an acronym for LInear Structural RELations) (K. G. Jöreskog, 1970, 1993; Jöreskog &

Sörbom, 1982; Jöreskog & Van Thillo, 1972; Jöreskog, Sorbom, & Magidson, 1979) and the AMOS (Analysis of Moment Structures) (Fox 2002, 2006; McArdle, 1988; McArdle & Epstein, 1987; McArdle & Hamagami, 1996, 2001). Variations on these algorithms have been implemented in EQS, TETRAD, and other packages.

Methods in systems of equation modeling and social network analytics are not as familiar in the social sciences as the first two methods, but offer comparatively more analytical power. Accessible and comprehensive tools for these additional approaches are covered in this book, as are research approaches to take advantage of the additional explanatory power that these approaches offer to social science research.

The breadth of application of SEM methods has been expanding, with SEM increasingly applied to exploratory, confirmatory, and predictive analysis through a variety of ad hoc topics and models. SEM is particularly useful in the social sciences where many if not most key concepts are not directly observable, and models that inherently estimate latent variables are desirable. Because many key concepts in the social sciences are inherently latent, questions of construct validity and methodological soundness take on a particular urgency. The popularity of SEM path analysis methods in the social sciences in one sense reflects a more holistic, and less blatantly causal, interpretation of many real-world phenomena than is found in the natural sciences. Direction in the directed network models of SEM arises from presumed cause-effect assumptions made about reality. Social interactions and artifacts are often epiphenomena—secondary phenomena that are difficult to directly link to causal factors. An example of a physiological epiphenomenon is, for example, time to complete a 100-m sprint. I may be able to improve my sprint speed from 12 to 11 s—but I will have difficulty attributing that improvement to any direct causal factors, like diet, attitude, and weather. The 1-s improvement in sprint time is an epiphenomenon—the holistic product of interaction of many individual factors. Such epiphenomena lie at the core of many sociological and psychological theories, and yet are impossible to measure directly. SEM provides one pathway to quantify concepts and theories that previously had only existed in the realm of ideological disputations.

To this day, methodologies for assessing suitable sample size requirements are a worrisome question in SEM-based studies. The number of degrees of freedom in structural model estimation increases with the number of potential combinations of latent variables, while the information supplied in estimating increases with the number of measured parameters (i.e., indicators) times the number of observations (i.e., the sample size)—both are nonlinear in model parameters. This should imply that requisite sample size is *not* a linear function solely of indicator count, even though such heuristics are widely invoked in justifying SEM sample size. Monte Carlo simulation in this field has lent support to the nonlinearity of sample size requirements. Sample size formulas for SEM are provided in the latter part of this book, along with assessments of existing rules of thumb. Contrary to much of the conventional wisdom, sample size for a particular model is constant across methods—PLS-PA, LISREL, and systems of regression approaches all require similar sample

sizes when similar models are tested at similar power and significance levels. None of these methods generates information that is not in the sample, though at the margin particular methods may more efficiently use sample information in specific situations.

1.1 The Problem Domains of Structural Equation Models

Many real-world phenomena involve networks of theoretical constructs of interest to both the natural and social sciences. Structural equation modeling has evolved to help specify real-world network models to fit observations to theory. Early approaches lacked the computational power to do little more than trace out pathways along the networks under study. These so-called path models were initially applied in the natural sciences to map networks of heritable genetic traits: constructs such as black hair, long ears, and so forth in laboratory animals, with relationship links defined by ancestry. This early research was directed towards developing useful models of inheritance from straightforward observation, without the benefit of pre-existing theories. In case a model did not at first fit the data, researchers had easy access to additional observations that theoretically could be replicated without end, simply by breeding another generation.

Quantification of the social sciences during the mid-twentieth century demanded statistical methods that could assess the abstract and often unobservable constructs inherent in these softer research areas. Many social science observations, e.g., a year of economic performance in the US economy, could never be repeated or replicated. Data was the product of quasi-experiments: non-replicable, with potential biases controlled via expanded scope rather than replication of the experiment. Early work empirically tested pairwise relationships between soft constructs using canonical correlations (Hotelling, 1936). Structural equation methods for the social sciences grew out of the empirical quantification of social research constructs, with pioneering work by Rensis Likert in psychology, politics, and sociology (R. Likert, 1932; Rensis Likert, Roslow, & Murphy, 1934; Quasha & Likert, 1937), Edward Deming (Stephan, Deming, & Hansen, 1940) in national census statistics and Lee Cronbach in education (Cronbach & Meehl, 1955). Pursuit of network models in the social sciences fostered the development of SEM models paralleling economic's systems of regression approaches, but supporting the empirical assessment of networks of unobservable constructs. These SEM approaches (PLS-PA and LISREL) evolved in the social sciences from statistical methods in canonical correlation, and were designed to fit data to networks of unobservable constructs. Herman Wold (1966, 1973, 1974, 1975) and his student, former high school teacher Karl Jöreskog (K. G. Jöreskog & Van Thillo, 1972), were focused on hypothesis testing simple theories about the structural relationships between unobservable quantities in social sciences. Applications started with economics, but found greater usefulness in measuring unobservable model constructs such as intelligence, trust, value, and so forth in psychology, sociology, and consumer sentiment. These approaches remain popular

today, as many central questions in the social sciences involve networks of abstract ideas that are often hard to measure directly.

Studies in economics and finance were less inclined towards abstractions, rather were challenged with the analysis and understanding of masses of dollar-denominated financial data. In this more tangible world, systems of regression models borrowed heavily from methods for analysis of transportation networks. These were employed to expand and extend existing linear regression modeling approaches (which themselves had been coopted from the astronomers). Systems of regression equation approaches pioneered by Tjalling Koopmans (1951, 1957, 1963) were designed to *prove or disprove theories about the structural relationships between economic measurements*. Because regression fits actual observations rather than abstract concepts, they are able to provide a wealth of goodness-of-fit information to assess the quality of the theoretical models tested. Wold and Jöreskog's methods provide goodness-of-fit information that in comparison is sparse, unreliable, and difficult to interpret.

Over the past two decades, many computer-intensive approaches have been developed to specify network models from data and to simulate observed behavior of real-world networks. Social network analysis extends graph theory into empirical studies. It takes observations (e.g., Wright's genetic traits) and classifies them as “nodes” (also called “vertices”). It infers relationships, called “edges” or “links” between these nodes from real-world observations. Social networks reflect social relationships in terms of individuals (nodes) and relationships (links) between the individuals. Examples of links are contracts, acquaintances, kinship, employment, and romantic relationships.

1.2 Motivation and Structure of This Book

The performance and behavioral characteristics of the three main structural equation model approaches have been reasonably well understood since their inception in the 1960–1970s. Unfortunately, for whatever reason, their strengths and weaknesses are too often misrepresented, or worse obscured behind sometimes specious fit statistics. The remainder of this book is designed to help the reader better understand what is happening on “the path”—specifically, how should a path coefficient be interpreted in a structural equation model. I try to keep things simple—my examples have purposely been limited to the simplest models possible, typically involving only two latent variables and one path. I also offer a comprehensive and accessible review of the current state of the art that will allow researchers to maximize research output from their datasets and confidently use available software for data analysis. The reader will find that the mathematics have purposely been kept to a minimum. The availability of extensive computer software, many with well-developed graphical interfaces, means that researchers can focus their efforts on assuring that research questions, design, data collection, and interpretation are accurate and complete while leaving the technical intricacies of data selection, tabulation, and statistical calculation to the software.

This book itself came together in small parts over a period of about 7 years. During those years—in my editorial roles at a number of journals—I reviewed many papers invoking path analysis methods, and repeatedly saw the same errors. Many of these problems arise because SEM software is now so well developed. Modern SEM software makes it possible to crank out results for complex, networked models that are based on only the vaguest of intuition, and then to test these same intuitions with data that does not directly measure them—and to do it all without understanding the statistics. In the wrong hands, this is surely a recipe for bad science. Though I will not dwell on research failure, I will occasionally cite cases where misuse of SEM has generated bad science, and what the consequences of these failures have been.

I hope that this book can provide a guide to generating reliable, defensible, and publishable answers to research questions, without the researcher expending an inordinate amount of time and effort trying to understand the philosophical debates and mathematical intricacies that allow us to confidently apply the techniques of SEM path analysis. In that sense, this book treats SEM as a “black box”—a box festooned with cranks, dials, knobs, and input readers. The researcher needs to know the consequences of particular controls and settings, and the meaning of particular dials, but need not understand the inner workings of the box. When driving a car, it is safer to spend all one’s attention making corrections for road situations than to worry about engine timing; the analogy carries over to SEM path analysis.

Chapter 2 surveys the development of structural equation models—from its roots in genetics, and the concessions made to accommodate the labor-intensive manual computation of the day, to the computationally intensive approaches that proliferate today. I attempt to accurately portray the trade-offs and controversies that each approach engendered along the way.

Chapter 3 traces the statistical debates that evolved in the 1950s as the USA and the World were seeking out national economic and social statistics, and the development of early partial least square path analysis (PLS-PA) methods with canonical correlations. Prior to WWII there were few if any national statistics to guide policy making at a national level. The USA experienced an unprecedented period of ferment in conception and development of social and economic statistics in the late 1940–1950s (Karabell, 2014). Challenged to find new ways to substantiate the abstract ideas that drove spending on social security, education, banking, and numerous other government programs, statisticians derived latent variable statistical approaches, approaches that fit whole networks of metrics, and methods that took advantage of the rapidly developing computing industry.

Chapter 4 surveys one offshoot of canonical correlation network analysis called LInear Structural RELations or LISREL for short. There have been many variations on LISREL (AMOS, EQS, TETRAD) providing the same results with slightly different algorithms. Chapter 5 looks at another offshoot of network analysis that added linear regression (originally developed in astronomy) to analytics for transportation networks. These systems of regression equation approaches became standard analytical tools in the economics community.

Chapters 6 and 7 explain the role of data in model building, testing, and inference, and offer sample size formulas and explanations of procedures for differing types and distributions of data. Chapter 6 covers the standard parametric distributions, and Chap. 7 focuses on Likert scale survey data to which SEM is commonly applied.

Chapter 8 explores the many aims that would motivate a researcher to use SEM in addressing a research question. Since many questions in social sciences can only be addressed through individual perceptions, impressions, and judgments, questions of objective truth are generally somewhat elusive. Questions involving a consumer's willingness to pay for a product or service, for example, can be measured prior to purchase by the stated buying intentions of a group of consumers; or they could be measured after purchase by the money spent: both are noisy signals about "willingness." Correctly and reliably answering research questions give statisticians insight into the "true state of nature," into the real world, and into truth.

Chapter 9 surveys the emerging successors to structural equation modeling in the social sciences that are classified generally under the rubric of social network analysis. The paths of structural equation models were always a kludge, a hodgepodge of early twentieth-century statistical techniques cobbled together to help explore naturally occurring networks. Scientists in the past simply did not possess the analytical tools to map more than a few links at a time; path models were the best we could do at the time. But modern approaches are harnessing the power of advanced processors, cloud storage, and sophisticated algorithms to finally make possible research paradigms about which turn of the twentieth-century researchers could only dream.

The remainder of this book is designed to help the reader better understand what is happening on "the path"—specifically, how should a path coefficient be interpreted in a structural equation model. I try to keep things simple—my examples have purposely been limited to the simplest models possible, typically involving only two latent variables and one path. I also offer a comprehensive and accessible review of the current state of the art that will allow researchers to maximize research output from their datasets and confidently use available software for data analysis. The reader will find that the mathematics have purposely been kept to a minimum. The availability of extensive computer software, many with well-developed graphical interfaces, means that researchers focus their efforts on assuring that research questions, design, data collection, and interpretation are accurate and complete while leaving the technical intricacies of data selection, tabulation, and statistical calculation to the software.

I hope that this book can provide a guide to generating reliable, defensible, and publishable answers to research questions, without the researcher expending an inordinate amount of time and effort trying to understand the philosophical debates and mathematical intricacies that allow us to confidently apply the techniques of SEM path analysis. The researcher needs to know the consequences of particular controls and settings, and the meaning of particular dials, but need not understand the inner workings of the box. By the end of this book, aspiring researchers should have a very good grasp of the knobs, dials, and controls available in both existing and emerging methods in SEM and network analysis.

References

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281.

Fox, J. (2006). Structural equation modeling with the SEM package in R. *Structural Equation Modeling*, 13(3), 465–486.

Fox, J. (2002). Structural equation models. *CRAN website*.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, 28(3/4), 321–377.

Jöreskog, K. G. (1970). A general method for estimating a linear structural equation system. *ETS Research Bulletin Series*, 1970, i–41.

Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Sage focus editions* (Vol. 154, p. 294). Thousand Oaks, CA: Sage.

Jöreskog, K. G., & Sörbom, D. (1982). Recent developments in structural equation modeling. *Journal of Marketing Research*, 19, 404–416.

Jöreskog, K. G., Sörbom, D., & Magidson, J. (1979). *Advances in factor analysis and structural equation models*. Cambridge, MA: Abt books.

Jöreskog, K. G., & Van Thillo, M. (1972). LISREL: a general computer program for estimating a linear structural equation system involving multiple indicators of unmeasured variables. *ETS Research Bulletin Series*, 1972, i–72.

Karabell, Z. (2014). *The leading indicators: a short history of the numbers that rule our world*. New York, NY: Simon and Schuster.

Koopmans, T. C. (1951). Analysis of production as an efficient combination of activities. In *Activity analysis of production and allocation* (Vol. 13, pp. 33–37).

Koopmans, T. C. (1957). *Three essays on the state of economic science* (Vol. 21). New York, NY: McGraw-Hill.

Koopmans, T. C. (1963). *Appendix to 'On the Concept of Optimal Economic Growth': Cowles Foundation for Research in Economics*. New Haven, CT: Yale University.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22, 1–55.

Likert, R., Roslow, S., & Murphy, G. (1934). A simple and reliable method of scoring the Thurstone attitude scales. *The Journal of Social Psychology*, 5(2), 228–238.

Lohmöller, J.-B. (1988). The PLS program system: latent variables path analysis with partial least squares estimation. *Multivariate Behavioral Research*, 23(1), 125–127.

Lohmöller, J.-B. (1989). *Latent variable path modeling with partial least squares*. Heidelberg: Physica-Verlag.

McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (pp. 561–614). New York, NY: Springer.

McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, 58, 110–133.

McArdle, J. J., & Hamagami, F. (1996). Multilevel models from a multiple group structural equation perspective. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: issues and techniques* (pp. 89–124). Mahwah, NJ: Erlbaum.

McArdle, J. J., & Hamagami, F. (2001). Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change decade of behavior* (pp. 139–175). Washington, DC: US American Psychological Association.

Quasha, W. H., & Likert, R. (1937). The revised Minnesota paper form board test. *Journal of Education and Psychology*, 28(3), 197.

Stephan, F. F., Deming, W. E., & Hansen, M. H. (1940). The sampling procedure of the 1940 population census. *Journal of the American Statistical Association*, 35(212), 615–630.

Wold, H. (1966). Estimation of principal components and related models by iterative least squares. *Multivariate Analysis*, 1, 391–420.

Wold, H. (1973). Nonlinear iterative partial least squares (NIPALS) modelling: some current developments. *Multivariate Analysis*, 3, 383–407.

Wold, H. (1974). Causal flows with latent variables: partings of the ways in the light of NIPALS modelling. *European Economic Review*, 5(1), 67–86.

Wold, H. (1975). *Path models with latent variables: the NIPALS approach*. New York, NY: Academic Press.

Chapter 2

A Brief History of Structural Equation Models

Though structural equation models today are usually associated with soft problems in the social sciences, they had their origin in the natural sciences—specifically biology. Europe’s nineteenth-century scholars were challenged to make sense of the diverse morphologies observed during an age of explorations, in Asia, Africa, and the Americas, as well as at home. In this period, new species of plants and animals were transplanted, domesticated, eaten, and bred at an unprecedented rate. An American ultimately provided one statistical tool that allowed scholars to build a science out of their diverse observations.

2.1 Path Analysis in Genetics

Seldom has a nonhuman animal been so thoroughly poked, observed, trained, and dissected as the domesticated dog. A member of the *Canidae* family, the dog is distantly related to coyotes and jackals, dingoes and dholes, foxes and wolves. There is evidence of distinct dog breeds as early as 5,000 years ago in drawings from ancient Egypt. The business of designing dogs for particular purposes began in earnest around the sixteenth century, and by the nineteenth century, clubs and competitions abounded for the naming and monitoring of breeds. There is a huge variation of sizes, shapes, temperaments, and abilities in modern dogs—much more so than in their homogeneous wolf ancestors. This has resulted from humans consciously influencing the genetics of dog populations through an involved network of interbreeding and active selection.

But none of this was a science at the dawn of the twentieth century, despite enormous expenditures, and centuries of breeding and contests to create “the perfect dog.” There was no theory (or perhaps too many competing but unsupported theories) about how particular characteristics arose in a particular subpopulation of dogs.