

Causation, Prediction, and Search

second edition



**Peter Spirtes,
Clark Glymour, and
Richard Scheines**

Causation, Prediction, and Search

Adaptive Computation and Machine Learning

Thomas Dietterich, Editor

Christopher Bishop, David Heckerman, Michael Jordan, and Michael Kearns,
Associate Editors

Bioinformatics: The Machine Learning Approach

Pierre Baldi and Søren Brunak

Reinforcement Learning: An Introduction

Richard S. Sutton and Andrew G. Barto

Graphical Models for Machine Learning and Digital Communication

Brendan J. Frey

Learning in Graphical Models

Michael I. Jordan

Causation, Prediction, and Search, second edition

Peter Spirtes, Clark Glymour, and Richard Scheines

Causation, Prediction, and Search

Peter Spirtes, Clark Glymour, and Richard Scheines

with additional material by

David Heckerman, Christopher Meek,
Gregory F. Cooper, and Thomas Richardson

The MIT Press
Cambridge, Massachusetts
London, England

©2000 Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Spirtes, Peter.

Causation, prediction, and search.—2nd ed. / Peter Spirtes, Clark Glymour, and Richard Scheines ; with additional material by David Heckerman, Christopher Meek, Gregory F. Cooper, and Thomas Richardson.

p. cm. — (Adaptive computation and machine learning)

Includes bibliographical references and index.

ISBN 0-262-19440-6 (hc : alk. paper)

I. Mathematical statistics. I. Glymour, Clark N. II. Scheines, Richard.
III. Title. IV. Series.

QA276 .S65 2000

519.—dc21

00-026266

Contents

Preface to the Second Edition	ix
Preface	xi
Acknowledgments	xv
Notational Conventions	xvii
1 Introduction and Advertisement	1
2 Formal Preliminaries	5
3 Causation and Prediction: Axioms and Explications	19
4 Statistical Indistinguishability	59
5 Discovery Algorithms for Causally Sufficient Structures	73
6 Discovery Algorithms without Causal Sufficiency	123
7 Prediction	157
8 Regression, Causation, and Prediction	191
9 The Design of Empirical Studies	209
10 The Structure of the Unobserved	253
11 Elaborating Linear Theories with Unmeasured Variables	269
12 Prequels and Sequels	295
13 Proofs of Theorems	377
Notes	475
Glossary	481
References	495
Index	531

To my parents, Morris and Cecile Spirtes—P.S.

In memory of Lucille Lynch Schwartz Watkins Speede Tindall Preston—C. G.

To Martha, for her support and love—R.S.

It is with data affected by numerous causes that Statistics is mainly concerned. Experiment seeks to disentangle a complex of causes by removing all but one of them, or rather by concentrating on the study of one and reducing the others, as far as circumstances permit, to comparatively small residuum. Statistics, denied this resource, must accept for analysis data subject to the influence of a host of causes, and must try to discover from the data themselves which causes are the important ones and how much of the observed effect is due to the operation of each.

—G. U. Yule and M. G. Kendall, 1950

The Theory of Estimation discusses the principles upon which observational data may be used to estimate, or to throw light upon the values of theoretical quantities, not known numerically, which enter into our specification of the causal system operating.

—Sir Ronald Fisher, 1956

George Box has [almost] said “The only way to find out what will happen when a complex system is disturbed is to disturb the system, not merely to observe it passively.” These words of caution about “natural experiments” are uncomfortably strong. Yet in today’s world we see no alternative to accepting them as, if anything, too weak.

—G. Mosteller and J. Tukey, 1977

Causal inference is one of the most important, most subtle, and most neglected of all the problems of Statistics.

—P. Dawid, 1979

Preface to the Second Edition

This second edition of *Causation, Prediction, and Search* is the culmination of almost twenty years of research on automation and causal inference, beginning in 1980 with a chapter of Glymour's *Theory and Evidence*, continuing with our book, *Discovering Causal Structure*, written with Kevin Kelly in 1987, and with an essay in 1990 (Spirtes et al. 1990) which laid out much of the research project we have followed in subsequent years. The thought—which one of us had—that the subject was more or less exhausted in 1993, when the first edition of this book appeared, has been proved entirely wrong.

For this edition we have substituted a new and briefer introduction, a discussion of d-separation that eliminates a misleading didacticism of the first edition, and an entirely new twelfth chapter, surveying and summarizing relevant results and applications since 1993. The original twelfth chapter was chiefly a series of conjectures, most of which have been proved correct, concerning cyclic graphs and feedback systems.

Our first debt for this edition is to our two former students, Chris Meek and Thomas Richardson. Much of the new work we describe is theirs. We are almost equally indebted to Gregory Cooper, David Heckerman, and Larry Wasserman, who have been wonderful, helpful colleagues and collaborators, and to Jaimie Robins, who, though often unhappy with the very idea of this book, helped with his insightfulness and fairness of mind. We have been encouraged by Judea Pearl's support, by his development of ideas presented here, particularly those on prediction first presented in chapter 7 of this book, and by his explorations of a multitude of new aspects of causal inference not considered here. We have been equally encouraged by the ingenious uses and modifications of our procedures provided by a number of scientists, including Bill Shipley, David Bessler and his collaborators, and Ludwig Litzka and his students. We owe a particular thanks to Cooper, Heckerman and Meek for permitting us to use in chapter 12 their survey of Bayesian search methods, and to Thomas Richardson for providing us with information about recent unpublished developments on chain graphs.

Preface

This book is intended for anyone, regardless of discipline, who is interested in the use of statistical methods to help obtain scientific explanations or to predict the outcomes of actions, experiments or policies.

Much of G. Udny Yule's work illustrates a vision of statistics whose goal is to investigate when and how causal influences may be reliably inferred, and their comparative strengths estimated, from statistical samples. Yule's enterprise has been largely replaced by Ronald Fisher's conception, in which there is a fundamental cleavage between experimental and non-experimental inquiry, and statistics is largely unable to aid in causal inference without randomized experimental trials. Every now and then members of the statistical community express misgivings about this turn of events, and, in our view, rightly so. Our work represents a return to something like Yule's conception of the enterprise of theoretical statistics and its potential practical benefits.

If intellectual history in the twentieth century had gone otherwise, there might have been a discipline to which our work belongs. As it happens, there is not. We develop material that belongs to statistics, to computer science, and to philosophy; the combination may not be entirely satisfactory for specialists in any of these subjects. We hope it is nonetheless satisfactory for its purpose. We are not statisticians by training or by association, and perhaps for that reason we tend to look at issues differently, and, from the perspective common in the discipline, no doubt oddly. We are struck by the fact that in the social and behavioral sciences, epidemiology, economics, market research, engineering, and even applied physics, statistical methods are routinely used to justify causal inferences from data not obtained from randomized experiments, and sample statistics are used to predict the effects of policies, manipulations, or experiments. Without these uses the profession of statistics would be a far smaller business. It may not strike many professional statisticians as particularly odd that the discipline thriving from such uses assures its audience that they are unwarranted, but it strikes us as very odd indeed. From our perspective outside the discipline, the most urgent questions about the application of statistics to such ends concern the conditions under which causal inferences and predictions of the effects of manipulations can and cannot reliably be made, and the most urgent need is a principled, rigorous theory with which to address these problems. To judge from the testimony of their books, a good many statisticians think any such theory is impossible. We think the common arguments against the possibility of inferring causes from statistics outside of experimental trials are unsound, and radical separations of the principles of experimental and observational study designs are unwise. Experimental and observational design may not always permit the same inferences, but they are subject to uniform principles.

The theory we develop follows necessarily from assumptions laid down in the statistical community over the last fifteen years. The underlying structure of the theory is essentially axiomatic. We will give two independent axioms on the relation between

causal structures and probability distributions and deduce from them features of causal relationships and predictions that can and that cannot be reliably inferred from statistical constraints under a variety of background assumptions. Versions of all of the axioms can be found in papers by Lauritzen, Wermuth, Speed, Pearl, Rubin, Pratt, Schlaifer, and others. In most cases we will develop the theory in terms of probability distributions that can be thought of loosely as propensities that determine long run frequencies, but many of the probability distributions can alternatively be understood as (normative) subjective degrees of belief, and we will occasionally note Bayesian applications. From the axioms there follow a variety of theorems concerning estimation, sampling, latent variable existence and structure, regression, indistinguishability relations, experimental design, prediction, Simpson's paradox, and other topics. Foremost among the "other topics" are the discovery that statistical methods commonly used for causal inference are radically suboptimal, and that there exist asymptotically reliable, computationally efficient search procedures that conjecture causal relationships from the outcomes of statistical decisions made on the basis of sample data. (The procedures we will describe require statistical decisions about the independence of random variables; when we say such a procedure is "asymptotically reliable" we mean it provides correct information if the outcome of each of the requisite statistical decisions is true in the population under study.)

This much of the book is mathematics: where the axioms are accepted, so must the theorems be, including the existence of search procedures. The procedures we describe are applicable to both linear and discrete data and can be feasibly applied to a hundred or more variables so long as the causal relations between the variables are sufficiently sparse and the sample sufficiently large. These procedures have been implemented in a computer program, TETRAD II, which at the time of writing is publicly available.¹

The theorems concerning the existence and properties of reliable discovery procedures of themselves tell us nothing about the reliabilities of the search procedures in the short run. The methods we describe require an unpredictable sequence of statistical decisions, which we have implemented as hypothesis tests. As is usual in such cases, in small samples the conventional p values of the individual tests may not provide good estimates of type 1 error probabilities for the search methods. We provide the results of extensive tests of various procedures on simulated data using Monte Carlo methods, and these tests give considerable evidence about reliability under the conditions of the simulations. The simulations illustrate an easy method for estimating the probabilities of error for any of the search methods we describe. The book also contains studies of one large pseudoempirical data set—a body of simulated data created by medical researchers to model emergency medicine diagnostic indicators and their causes—and a great many empirical data sets, most of which have been discussed by other authors in the context of specification searches.

A further aim of this work is to show that a proper understanding of the relationship between causality and probability can help to clarify diverse topics in the statistical literature, including the comparative power of experimentation versus observation,

Simpson's paradox, errors in regression models, retrospective versus prospective sampling, the perils of variable selection, and other topics. There are a number of relevant topics we do not consider. They include problems of estimation with discrete latent variables, optimizing statistical decisions, many details of sampling designs, time series, and a full theory of "nonrecursive" causal structures—that is, finite graphical representations of systems with feedback.

Causation, Prediction, and Search is not intended to be a textbook, and it is not fitted out with the associated paraphernalia. There are open problems but no exercises. In a textbook everything ought to be presented as if it were complete and tidy, even if it isn't. We make no such pretenses in this book, and the chapters are rich in unsolved problems and open questions. Textbooks don't usually pause much to argue points of view; we pause quite a lot.

The various theorems in this book often have a graph theoretic character; many of them are long, difficult case arguments of a kind quite unfamiliar in statistics. In order not to interrupt the flow of the discussion we have placed all proofs but one in a chapter at the end of the book. In the few cases where detailed proofs are available in the published literature, we have simply referred the reader to them. Where proofs of important results have not been published or are not readily available we have given the demonstrations in some detail.

The structure of the book is as follows. Chapter 1 concerns the motivation for the book in the context of current statistical practice and advertises some of the results. Chapter 2 introduces the mathematical ideas necessary to the investigation, and chapter 3 gives the formal framework a causal interpretation, lays down the axioms, notes circumstances in which they are likely to fail, and provides a few fundamental theorems. The next two chapters work out the consequences of two of the axioms for some fundamental issues in contexts in which it is known, or assumed, that there are no unmeasured common causes affecting measured variables. In chapter 4 we give graphical characterizations of necessary and sufficient conditions for causal hypotheses to be statistically indistinguishable from one another in each of several senses. In chapter 5 we criticize features of model specification procedures commonly recommended in statistics, and we describe feasible algorithms that from properties of population distributions extract correct information about causal structure, assuming the axioms apply and that no unmeasured common causes are at work. The algorithms are illustrated for a variety of empirical and simulated samples. Chapter 6 extends the analysis of chapter 5 to contexts in which it cannot be assumed that no unmeasured common causes act on measured variables. From both a theoretical and practical perspective, this chapter and the next form the center of the book, but they are especially difficult. Chapter 7 addresses the fundamental issue of predicting the effects of manipulations, policies, or experiments. As an easy corollary, the chapter unifies directed graphical models with Donald Rubin's "counterfactual" framework for analyzing prediction. Chapter 8 applies the results of the preceding chapters to the subject of regression. We argue that even

when standard statistical assumptions are satisfied multiple regression is a defective and unreliable way to assess causal influence even in the large sample limit, and various automated regression model specification searches only make matters worse. We show that the algorithms of chapter 6 are more reliable in principle, and we compare the performances of these algorithms against various multiple regression procedures on a variety of simulated and empirical data sets. Chapter 9 considers the design of empirical studies in the light of the results of earlier chapters, including issues of retrospective and prospective sampling, the comparative power of experimental and observational designs, selection of variables, and the design of ethical clinical trials. The chapter concludes with a look back at some aspects of the dispute over smoking and lung cancer. Chapters 10 and 11 further consider the linear case, and analyze algorithms for discovering or elaborating causal relations among measured and unmeasured variables in linear systems. Chapter 12 is a brief consideration of a variety of open questions. Proofs are given in chapter 13.

We have tried to make this work self-contained, but it is admittedly and unavoidably difficult. The reader will be aided by a previous reading of Pearl 1988, Whittaker 1990, or Neapolitan 1990.

Acknowledgments

One source of the ideas in this book is in work we began ten years ago at the University of Pittsburgh. We drew many ideas about causality, statistics, and search from the psychometric, economic, and sociological literature, beginning with Charles Spearman's project at the turn of the century and including the work of Herbert Simon, Hubert Blalock, and Herbert Costner.

We obtained a new perspective on the enterprise from Judea Pearl's *Probabilistic Reasoning in Intelligent Systems*, which appeared the next year. Although not principally concerned with discovery, Pearl's book showed us how to connect conditional independence with causal structure quite generally, and that connection proved essential to establishing general, reliable discovery procedures. We have since profited from correspondence and conversation with Pearl and with Dan Geiger and Thomas Verma, and from several of their papers. Pearl's work drew on the papers of Wermuth (1980), Kiiveri and Speed (1982), Wermuth and Lauritzen (1983), and Kiiveri, Speed, and Carlin (1984), which in the early 1980s had already provided the foundations for a rigorous study of causal inference. Paul Holland introduced one of us to the Rubin framework some years ago, but we only recently realized it's logical connections with directed graphical models. We were further helped by J. Whittaker's (1990) excellent account of the properties of undirected graphical models.

We have learned a great deal from Gregory Cooper at the University of Pittsburgh who provided us with data, comments, Bayesian algorithms and the picture and description of the ALARM network which we consider in several places. Over the years we have learned useful things from Kenneth Bollen. Chris Meek provided essential help in obtaining an important theorem that derives various claims made by Rubin, Pratt, and Schlaifer from axioms on directed graphical models.

Steve Fienberg and several students from Carnegie Mellon's department of statistics joined with us in a seminar on graphical models from which we learned a great deal. We are indebted to him for his openness, intelligence, and helpfulness in our research, and to Elizabeth Slate for guiding us through several papers in the Rubin framework. We are obliged to Nancy Cartwright for her courteous but salient criticism of the approach taken in our previous book and continued here. Her comments prompted our work on parameters in chapter 4. We are indebted to Brian Skyrms for his interest and encouragement over many years, and to Marek Druzdzal for helpful comments and encouragement. We have also been helped by Linda Bouck, Ronald Christensen, Jan Callahan, David Papineau, John Earman, Dan Hausman, Joe Hill, Michael Meyer, Teddy Seidenfeld, Dana Scott, Jay Kadane, Steven Klepper, Herb Simon, Peter Slezak, Steve Sorensen, John Worrall, and Andrea Woody. We are indebted to Ernest Seneca for putting us in contact with Dr. Rick Linthurst, and we are especially grateful to Dr. Linthurst for making his doctoral thesis available to us.

Our work has been supported by many institutions. They, and those who made decisions on their behalf, deserve our thanks. They include Carnegie Mellon University, the National Science Foundation programs in History and Philosophy of Science, in Economics, and in Knowledge and Database Systems, the Office of Naval Research, the Navy Personnel Research and Development Center, the John Simon Guggenheim Memorial Foundation, Susan Chipman, Stanley Collyer, Helen Gigley, Peter Machamer, Steve Sorensen, Teddy Seidenfeld, and Ron Overmann. The Navy Personnel Research and Development Center provided us the benefit of access to a number of challenging data analysis problems from which we have learned a great deal.

Notational Conventions

Text

In the text, each technical term is written in boldface where it is defined.

Variables:	capitalized, and in italics, e.g., X
Values of variables:	lower case, and in italics, e.g., $X = x$
Sets:	capitalized, and in boldface, e.g., \mathbf{V}
Values of sets of variables:	lower case, and in boldface, e.g., $\mathbf{V} = \mathbf{v}$
Members of \mathbf{X} that are not members of \mathbf{Y} :	$\mathbf{X} \setminus \mathbf{Y}$
Error variables:	ε, δ, e
Independence of \mathbf{X} and \mathbf{Y} :	$\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$
Independence of \mathbf{X} and \mathbf{Y} conditional on \mathbf{Z} :	$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$
$\mathbf{X} \cup \mathbf{Y}$:	\mathbf{XY}
Covariance of X and Y :	$\text{COV}(X, Y)$ or γ_{XY}
Correlation of X and Y :	ρ_{XY}
Sample correlation of X and Y :	r_{XY}
Partial Correlation of X and Y , controlling for all members of set \mathbf{Z} :	$\rho_{XY \cdot \mathbf{Z}}$

In all of the graphs that we consider, the vertices are random variables. Hence we use the terms “variables in a graph” and “vertices in a graph” interchangeably.

Figures

Figure numbers occur just below a figure, starting at 1 within each chapter. Where necessary, we distinguish between measured and unmeasured variables by boxing measured variables and circling unmeasured variables (except for error terms). Variables beginning with e , ε , or δ are understood to be “error,” or “disturbance,” variables. For example, in the figure below, X and Y are measured, T is not, and ε is an error term.

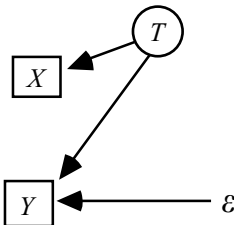


Figure n.1

We will neither box nor circle variables in graphs in which no distinction need be made between measured and unmeasured variables, for example, figure n.2.



Figure n.2

For simplicity, we state and prove our results for probability distributions over discrete random variables. However, under suitable integrability conditions, the results can be easily generalized to continuous distributions that have density functions by replacing the discrete variables by continuous variables, probability distributions by density functions, and summations by integrals.

If a description of a set of variables is a function of a graph G and variables in G , then we make G an optional argument to the function. For example, $\mathbf{Parents}(G, X)$ denotes the set of variables that are parents of X in graph G ; if the context makes clear which graph is being referred to we will simply write $\mathbf{Parents}(X)$.

If a distribution is defined over a set of random variables \mathbf{O} then we refer to the distribution as $P(\mathbf{O})$. An equation between distributions over random variables is understood to be true for all values of the random variables for which all of the distributions in the equation are defined. For example if X and Y each take the values 0 or 1 and $P(X=0) \neq 0$ and $P(X=1) \neq 0$ then $P(Y|X) = P(Y)$ means $P(Y=0|X=0) = P(Y=0)$, $P(Y=0|X=1) = P(Y=0)$, $P(Y=1|X=0) = P(Y=1)$, and $P(Y=1|X=1) = P(Y=1)$.

We sometimes use a special summation symbol, $\overset{\rightarrow}{\sum}$, which has the following properties:

- (i) when sets of random variables are written beneath the special summation symbol, it is understood that the summation is to be taken over sets of values of the random variables, not the random variables themselves,
- (ii) if a conditional probability distribution appears in the scope of such a summation symbol, the summation is to be taken only over values of the random variables for which the conditional probability distributions are defined,
- (iii) if there are no values of the random variables under the special summation symbol for which the conditional probability distributions in the scope of the symbol are defined, then the summation is equal to zero.

For example, suppose that X , Y , and Z can each take on the values 0 or 1. Then if $P(Y=0, Z=0) = 0$

$$\sum_X^{\rightarrow} P(X|Y=0, Z=0) = P(X=0|Y=0, Z=0) + P(X=1|Y=0, Z=0)$$

However, if $P(Y=0, Z=0) = 0$, then $P(X=0|Y=0, Z=0)$ and $P(X=1|Y=0, Z=0)$ are not defined, so

$$\sum_X^{\rightarrow} P(X|Y=0, Z=0) = 0$$

We will adopt the following conventions for empty sets of variables. If $\mathbf{Y} = \emptyset$ then

- (i) $P(\mathbf{X}|\mathbf{Y})$ means $P(\mathbf{X})$.
- (ii) $\rho_{\mathbf{XZ},\mathbf{Y}}$ means $\rho_{\mathbf{XZ}}$.
- (iii) $\mathbf{A} \perp\!\!\!\perp \mathbf{B}|\mathbf{Y}$ means $\mathbf{A} \perp\!\!\!\perp \mathbf{B}$.
- (iv) $\mathbf{A} \perp\!\!\!\perp \mathbf{Y}$ is always true.

1 Introduction and Advertisement

1.1 The Issue

Adult judgments about which event is “the cause” of another event are loaded with topicality, interest, background knowledge about normal cases, and moral implications. Tell someone that Suzie was injured in an accident while John was driving her home, and then ask what further information is needed to decide whether John’s actions caused her injury. People want to know John’s condition, the detailed circumstances of the accident, including the condition of the roadway, of John’s car, of the other driver if there was one, and so on (Ahn 1995, 1996). The responses show that in such contexts judgments about causation have a moral aspect, and an aspect that depends on an understanding of normal conditions and deviations from the normal. That sort of thing will vary with culture, background, and circumstance.

Causal claims have a subjunctive complexity—they are associated with claims about what did not happen, or has not happened yet, or what would have happened if some circumstance had been otherwise. If someone says their hair is brown because they dye it, we infer that if they had not dyed their hair it would have been some other color. That sort of counterfactual conditional is not always correct (someone with brown hair can dye their hair brown), and endless but indispensable complexities result. Our moral sense, our very notions of blame and regret, depend on subjunctive aspects of causal claims. In addition, the kinds of entities that are described as causes and effects are enormously varied, and the logical form of causal claims can vary from particular to general to universal. Events are causes—*the rise of the middle class caused the American Revolution; Constantine’s conversion caused the triumph of Christianity in the Roman Empire; the discovery of penicillin saved millions of lives*. Features or properties, or their changes, are often cited as causes—*the pH of the liquid caused it to turn pink when phenolphthalein was added; the heat caused the butter to melt*. Objects or persons are cited as causes—*my daughter gave me a cold*. Even relationships, or instances of them, can be described as causes—*her love for him caused her to leave the country*. Descriptions of effects can be equally varied. The salient effect of a preventive cause, for example, can be a circumstance or event that doesn’t exist—*she prevented the catastrophe*.

The variation—the looseness—of causal claims has provided a reason for many people to dismiss the very idea of causation as prescientific. Bertrand Russell claimed as much, and Karl Pearson proposed to replace the idea of causation entirely by the idea of correlation. To this day, some writers try to avoid the issue by euphemism, as though a new word would clarify things, and at almost any conference of statisticians or social scientists (but not, anymore, of philosophers) there is someone—often not alone—who is eager to say he doesn’t “believe in causality.” But he acts as if he does; we all do, all the time: we ask people to do things, or do them ourselves, because we want what we think

will result from the actions—*turn down the volume on the radio, its too loud*—and we blame people for the unhappy effects of their actions. The skeptic about causality pushes the brake pedal to make his car slow, flips a switch to make a lamp glow, puts his money in the bank to collect interest.

Francis Bacon claimed knowledge is power, and he was talking about the power of control supplied by causal knowledge. One of the greatest mysteries of the human condition is that in a few short years a newborn infant comes to control much of her environment, knows how to climb up to things, how to turn on the television with the remote control, how to make a balloon expand and a soap bubble form, how to summon others and how to avoid them. Developmental psychology has hardly begun to crack how all that causal knowledge, all that power to control, is acquired so quickly. And a great deal—perhaps most—of our scientific inquiry aims to find out something about causal relationships. Billions are spent each year to discover the effects of drug treatments alone, and similar sums to estimate the likely results of possible social and economic policies. Those who claim not to believe in causality may provide consulting services to clients who want to predict the effects of alternative business strategies, or who want to know how to judge the bearing of some body of data on a causal hypothesis. Loose as the notion may be, there is nothing serious to the claim that we can live and thrive without using the idea of causation, however we name it.

The baby and the scientist occupy two ends of the same question: how can observations be turned into causal knowledge, and how can causal knowledge, even if incomplete, be used to influence and control our environment? The theory of experimental design offers a route to causal knowledge, but while Fisher's discussions of the probabilistic and statistical aspects of experimental design were brilliant and rigorous, everything causal was left informal. Fisher did not provide as rigorous a theory regarding causal inference from non-experimental observations. Yet most of what we want to know about, and most of what we think we know, is not amenable to randomized clinical trials.

The question of prediction has been equally unsettled—the question is: if you know some causal relations, and you know some of the probability relations among some of the related variables, can you predict what will result if you intervene and alter the value of one or more of the variables. In many causal systems the probability of an event Y given an intervention to bring about an event X is different from the conditional probability of Y on X . In recent years some philosophers, economists, computer scientists, and statisticians have realized the importance to many different kinds of problems of the difference between predicting by conditioning and predicting by intervening. Philosophers have used the difference between conditioning and intervening to argue that the principle of maximum expected utility is not always rational. Whether they are right or wrong about that (and we think wrong: see Meek and Glymour 1994), the essential thing is to provide a general means of determining when the probability distribution of one variable can be calculated from an intervention that forces a probability distribution on the values of another variable, given partial causal and probabilistic knowledge of the

undisturbed system, and, when the probability of an effect can be calculated, to provide a means of calculating it.

So we have three problems: first, the problem of clarifying the very idea of a causal system with sufficient precision for mathematical analysis and sufficient generality to capture a wide range of scientific practices; second, the problem of understanding the possibilities and limitations for discovering such causal structures from various kinds of data; and third, the problem of characterizing the probabilities predicted by a causal hypothesis given an intervention directly to force a value, or distribution of values, on one or more variables. This book attempts answers to all three of these questions.

Our answer to the problem of regimenting causal hypotheses uses a formalism developed by Terry Speed and his students and subsequently elaborated by Judea Pearl and his students, and gives it a causal interpretation previously suggested (Kiiveri and Speed 1982). Our answer to the problem of discovery turns on algorithms developed from the mathematics of the representation; that answer is supplemented in chapter 12 by a discussion of Bayesian algorithms lent us by David Heckerman, Christopher Meek, and Gregory Cooper, and by a discussion—based on joint work with Wasserman and Robins—of the convergence properties of any possible non-experimental discovery procedure. The assumptions of the theory of manipulation developed here has anticipations in the econometric literature (Strotz and Wold 1960), but by putting these assumptions in a graphical framework we are able to prove some novel theorems that follow from the assumptions.

One approach to clarifying the notion of causation—the philosophers' approach ever since Plato—is to try to define "causation" in other terms, to provide necessary and sufficient and noncircular conditions for one thing, or feature or event or circumstance, to cause another, the way one can define "bachelor" as "unmarried adult male human." Another approach to the same problem—the mathematician's approach ever since Euclid—is to provide axioms that use the notion of causation without defining it, and to investigate the necessary consequences of those assumptions. We have few fruitful examples of the first sort of clarification, but many of the second: Euclid's geometry, Newton's physics, Frege's logic and Hilbert's, Kolmogorov's probability. Some axiomatic theories—Newton's, for example—offer a substantive theory of nature, while others—Frege's and Hilbert's logics and Kolmogorov's probability—are a systematization and abstraction from practice and intuition—while still others—Euclid's *Elements*—are something of both. While we do not claim the success of these examples, they are the models of this book.

We use a formalism—directed graphical models—that is not in the least original with us; we claim some originality in explicitly stating the causal assumptions implicit in the causal interpretation of the graphs, and in extending the application of graphs to solving certain problems about manipulations. The representation invokes two ideas about causation that are fundamental and ancient. The first idea, which can be traced back at least to Bernoulli, is that the absence of causal relations is marked by independence in

probability—in Bernoulli’s examples, if the outcome of one trial has no influence on the outcome of another trial, then the probability of both outcomes equals the product of each outcome separately. The second idea, Bacon’s again, is that probability is associated with control: if variation of one feature, X , causes variation of another feature Y , then Y can be changed by an appropriate intervention that alters X . It turns out that the representation captures what is common to a wide variety of statistical models of causal relations—for example: regression models, logistic regression models, structural equation models, latent factor models, and many models of categorical data—and captures how these models may be used in prediction and control. The general assumptions are given in chapter 3.

These axioms have implications for scientific discovery by experimental and non-experimental means. Our investigations require characterizing when two or alternative causal theories are, in various technical senses, indistinguishable by data, and characterizing when and which causal features are shared by all models indistinguishable from any particular model. These characterizations are given in chapter 4, and more recent work on equivalence is described in chapter 12. In chapters 5, 6, 10, and 11, and again in chapter 12, we describe algorithms for discovering causal structure from sample data. We evaluate the algorithms in terms of several different features. (1) Are they computationally feasible on realistic problems? (2) Are they reliable—do they in some sense converge to a description of features common to all models indistinguishable from the true model? (3) Are they as informative as possible for the features of the data they use? Our algorithmic results concern the discovery of causal structure in linear and nonlinear systems, systems with and without feedback, cases where there may, or may not, be unrecorded common causes of recorded variables, and cases in which membership in the observed sample is influenced by the variables under study. We investigate the reliabilities of the algorithms on simulated data, and illustrate their application with published data sets. In chapter 8, reliable procedures for causal inference are compared with regression in theory, on simulated data, and on real data sets. In chapter 9 we compare causal inference from experimental and non-experimental data. Besides reviews of work on equivalence, prediction and search algorithms, chapter 12 includes a consideration of the senses in which it is, and is not, possible to have a procedure that reliably converges to the truth about causal relations as they are represented here, and a discussion of procedures for learning feedback models, so far as such models can be represented by directed cyclic graphs, and a brief discussion of the combination of search methods with the Gibbs sampler and related procedures for estimation of posterior probabilities.

The discovery of causal relations is only half of the story. The other half concerns the use in prediction of causal knowledge, even partial and incomplete causal knowledge. The fundamentals of the theory of prediction are given in chapter 3, and their consequences are developed in detail in chapter 7. The theory of prediction has a number of limitations, some of which are considered in chapter 12. The final chapter, 13, provides detailed proofs of the theorems in the main text.

References

- Ahn, W., Kalish, C., Medin, D., and Gelman, S. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition* 54: 299–352.
- Ahn, W., and Bailenson, J. (1996). Causal attribution as a search for underlying mechanisms: An explanation of the conjunction fallacy and the discounting principle. *Cognitive Psychology*.
- Aigner, D., and Goldberger, S. (1977). *Latent Variables in Socio-economic Models*. Amsterdam: North-Holland.
- Aitkin, M. (1979). A simultaneous test procedure for contingency table models. *Applied Statistics* 28: 233–242.
- Akleman, D., Bessler, D., and Burton, D. (1999). Modeling corn exports and exchange rates with directed graphs and statistical loss functions. In *Computation, Causation and Discovery*, edited by C. Glymour and G. Cooper. Cambridge, Mass.: MIT Press.
- Alchourrón, C., Gärdenfors, P., and Mackinson, C. (1985). On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic* 50: 510–530.
- Aliferis, C., and Cooper, G. (1994). An evaluation of an algorithm for inductive learning of Bayesian belief networks using simulated data sets. *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence*, Seattle: Morgan Kaufmann, 8–14.
- Allen, D. (1974). The relationship between variable selection and data augmentation and a method for prediction. *Technometrics* 17: 125–127.
- Anderson, J., and Gerbing, D. (1982). Some methods for respecifying measurement models to obtain unidimensional construct measurement. *Journal of Marketing Research* 19: 453–460.
- Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. New York: Wiley.
- Andersson, S., Madigan, D., and Perlman, M. (1996). An alternative Markov property for chain graphs. *Proceedings of the 12th Conference on Uncertainty in AI*, Portland, Ore.: Morgan Kaufmann, 40–48.
- Artzenius, F. (1992). The Common Cause Principle. *Philosophy of Science Association*: 227–237.

- Asher, Herbert B. (1976). *Causal Modeling*. Beverly Hills, Calif.: Sage Publications.
- Asmussen, S., and Edwards, D. (1983). Collapsibility and response variables in contingency tables. *Biometrika* 70: 567–578.
- Bagozzi, R. (1980). *Causal Models in Marketing*. New York: Wiley.
- Baldwin, B. (1986). The effects of structural model misspecification and sample size on the robustness of LISREL maximum likelihood parameter estimates. Department of Administrative and Foundational Services, Louisiana State University.
- Balke, A., and Pearl, J. (1994). Probabilistic Evaluation of Counterfactual Queries. *Proceedings of the Twelfth National Conference on Artificial Intelligence*. MIT Press, volume 1, 230–237.
- Bartlett, M. (1935). Contingency table interaction. *J. Roy. Statist. Soc. Suppl.* 2: 248–252.
- Bartlett, M. (1954). A note on the multiplying factors for various chi-squared approximations. *J. Roy. Statist. Soc. Ser. B* 16: 196–198.
- Basman, R. (1965). A note on the statistical testability of ‘explicit causal chains’ against the class of ‘interdependent’ models. *JASA* 60: 1080–1093.
- Beale, E., Kendall, M., and Mann, D. (1967). The discarding of variables in multivariate analysis. *Biometrika* 54: 357–366.
- Bearden, W. O., Sharma, S., and Teel, J. E. (1982). Sample size effects on chi-square and other statistics used in evaluating causal models. *Journal of Marketing Research* 19: 425–430.
- Becker, G. (1964). Human capital; a theoretical and empirical analysis, with special reference to education. New York: National Bureau of Economic Research; distributed by Columbia University Press.
- Becker, S., and LeCun, Y. (1989). Improving the convergence of back-propagation learning with second order methods. *Proceedings of the 1988 Connectionist Models Summer School*. Morgan Kaufmann.
- Beinlich, I., Suermondt, H., Chavez, R., and Cooper, G. (1989). The ALARM monitoring system: a case study with two probabilistic inference techniques for belief networks. *Proceedings of the Second European Conference on Artificial Intelligence in Medicine*. London: England.

- Bentler, P. (1980). Multivariate analysis with latent variables: causal modeling. *Annual Review of Psychology* 31: 419–456.
- Bentler, P. (1985). Theory and Implementation of EQS: A Structural Equations Program. Los Angeles, BMDP Statistical Software.
- Bentler, P. (1995). *EQS: Structural equations program manual (Version 5.0)*. Encino, Calif.: Multivariate Software.
- Bentler, P., and Bonett, D. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin* 88: 588–606.
- Bentler, P., and Peeler, W. (1979). Models of female orgasm. *Archives of Sexual Behavior* 8: 405–423.
- Bernardo, J., and Smith, A. (1994). *Bayesian Theory*. New York: Wiley.
- Bessler, D., and Akleman, D. (1998). Farm prices, retail prices and directed graphs: results for pork and beef. *American Journal of Agricultural Economics* 80: 1144–1149.
- Bessler, D., and Güven, D. (1997). A note on directed graphs and time series data. submitted to *Econometric Reviews*.
- Birch, M. (1963). Maximum likelihood in three-way contingency tables. *J. Roy. Statist. Soc.* 25: 220–223.
- Bishop, Y, Fienberg, S., and Holland, P. (1975). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, Mass.: MIT Press.
- Blalock, H. (1969). *Theory Construction; from Verbal to Mathematical Formulations*. Englewood Cliffs, N.J.: Prentice-Hall.
- Blalock, H. (1971). *Causal Models in the Social Sciences*. Chicago: Aldine-Atherton.
- Blalock, H. (1972). *Causal Inferences in Nonexperimental Research*. New York: Norton.
- Blau, P., and Duncan, O. (1967). *The American Occupational Structure*. New York: Wiley.
- Blum, R. (1984). Discovery, confirmation and incorporation of causal relationship from a time-oriented clinical data base: The RX project. *Readings in Medical Artificial Intelligence*, edited by W. Clancey and E. Shortliffe. Reading, Mass.: Addison-Wesley.

- Blyth, C. (1972). On Simpson's paradox and the sure-thing principal. *JASA* 67: 364–366.
- Bollen, K. (1990). Outlier screening and a distribution-free test for vanishing tetrads. *Sociological Methods and Research* 19: 80–92.
- Bollen, K. (1989). *Structural Equations with Latent Variables*. New York: Wiley.
- Boomsma, A. (1982). The robustness of LISREL against small sample sizes in factor analysis models. In *Systems Under Indirect Observation: Causality, Structure, Prediction (Part I)*, edited by K. Jöreskog and H. Wold. Amsterdam: North-Holland, 149–173.
- Boomsma, A. (1983). On the robustness of LISREL (maximum likelihood estimation) against small sample size and nonnormality. Amsterdam, Sociometric Research Foundation.
- Bowden, R., and Turkington, D. (1984). *Instrumental Variables*. Cambridge: Cambridge University Press.
- Boyer, X., Friedman, N., and Koller, D. (1999) Discovering the structure of complex dynamic systems. *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufman, 91–100.
- Breslow, N., and Day, N. (1980). *Statistical Methods in Cancer research, Vol. 1: The analysis of Case-Control Studies*. Lyon, IARC.
- Brownlee, K. (1965). A review of "Smoking and Health." *JASA*, 722–739.
- Bunker, J., Forrest, W., Mosteller, F., and Vandam, L. (1969). The National Halothane Study: Report of the Subcommittee on the National Halothane Study of the Committee on Anesthesia. Washington, D.C., Division of Medical Sciences, National Academy of Sciences, National Research Council.
- Buntine, W. (1991). Theory refinement on Bayesian networks. *Proceedings of the Seventh Conference on Uncertainty in AI*, Los Angeles, Calif.: Morgan Kaufmann, 52–61.
- Buntine, W. (1994). Operations for learning with graphical models. *Journal for Theoretical and Experimental Artificial Intelligence* 2: 159–225.
- Buntine, W. (1996). A guide to the literature on learning graphical models. *IEEE Transactions on Knowledge and Data Engineering* 8: 195–210.

- Burch, P. (1978). Smoking and lung cancer: The problem of inferring cause (with discussion). *J. Roy. Statist. Soc.* 141(Series A): 437–477.
- Burch, P. (1983). The Surgeon General's "Epidemiologic Criteria for Causality." A critique. *Journal of Chronic Diseases* 36: 821–836.
- Burch, P. (1984). The Surgeon General's "Epidemiologic Criteria for Causality." Reply to Lilienfeld. *Journal of Chronic Diseases* 37: 148–157.
- Byron, R. (1972). Testing for misspecification in econometric systems using full information. *International Economic Review* 28: 138–151.
- Callahan, J., and Sorenson, S. (1992). Using TETRAD II as an automated exploratory tool. *Social Science Computer Review* 10: 329–336.
- Campbell, D., Schwartz, R., Sechrest, L., and Webb, Eugene J. (1966). *Unobtrusive Measures; Nonreactive Research in the Social Sciences*. Chicago: Rand McNally.
- Campbell, D., Stanley, J., and Gage, N. (1966). *Experimental and quasi-experimental designs for research*. Chicago: R. McNally.
- Caramazza, A. (1986). On drawing inferences about the structure of normal cognitive processes from patterns of impaired performance: the case for single patient studies. *Brain and Cognition* 5: 41–66.
- Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford, New York: Clarendon Press; Oxford University Press.
- Cartwright, N. (1989). *Nature's Capacities and their Measurement*. Oxford, New York: Clarendon Press; Oxford University Press.
- Cartwright, N. (1993). Marks and probabilities: two ways to find a causal structure. In *Scientific Philosophy: Origins and Development*, edited by F. Stadler. Kluwer, Dordrecht.
- Casella, G., and George, E. (1992). Explaining the Gibbs sampler. *The American Statistician* 46: 167–174.
- Cavallo, R., and Klir, G. (1979). Reconstructability analysis of multi-dimensional relations: A theoretical basis for computer-aided determination of acceptable systems models. *International Journal of General Systems* 5: 143–171.

- Cederlof, R., Friberg, L., and Lundman, T. (1972). The interactions of smoking, environment and heredity and their implications for disease etiology. *Acta Med Scand.* 612 (Suppl.).
- Cheeseman, P., and Stutz, J. (1995). Bayesian classification (AutoClass): Theory and results. *Advances in Knowledge Discovery and Data Mining*, edited by U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, Menlo Park, Calif.: AAAI Press, 153–180.
- Chib, S. (1995). Marginal likelihood from the Gibbs output. *JASA* 90: 1313–1321.
- Chib, S., and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician* 49: 327–335.
- Chickering, D. (1996). Learning Bayesian networks is NP-complete. *Learning from Data*, edited by D. Lenz and H. Fisher. Springer-Verlag, 121–130.
- Chickering, D. (1996). Learning equivalence classes of Bayesian-network structures. *Proceedings of the Twelfth Conference on Uncertainty in AI*, Portland, Ore.: Morgan Kaufmann, 150–157.
- Chickering, D., and Heckerman, D. (1997). Efficient approximations for the marginal likelihood of Bayesian networks with hidden variables. *Machine Learning* 29: 181–212.
- Chou, C., Bentler, P., and Satorra, A. (1991). Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: A Monte Carlo study. *British Journal of Mathematical and Statistical Psychology* 85: 398–409.
- Chow, C., and Liu, C. (1968). Approximating discrete probability distributions with dependence trees. *IEEE Trans. on Info. Theory* IT-14: 462–467.
- Chow, C., and Wagner, T. (1973). Consistency of an estimate of tree-dependent probability distributions. *IEEE Trans. on Info. Theory* IT-19: 369–371.
- Christensen, R. (1990). *Log-linear Models*. New York: Springer-Verlag.
- Coleman, J. (1964). *Introduction to Mathematical Sociology*. New York: Free Press of Glencoe.
- Cooper, G. (1989). Current research in the development of expert systems based on belief networks. *Applied Stochastic Models and Data Analysis* 5: 39–52.

- Cooper, G. (1995). Causal discovery from data in the presence of selection bias. *Proceedings of the Fifth International Workshop on Artificial Intelligence and Statistics*, Fort Lauderdale, FL.
- Cooper, G. (1997). A simple constraint-based algorithm for efficiently mining causal observational databases for causal relationships. *Data Mining and Knowledge Discovery* 1: 201–224.
- Cooper, G., and Herskovits, E. (1991). A Bayesian method for constructing Bayesian belief networks from databases. *Proceedings of the Seventh Annual Conference on Uncertainty in AI*, Los Angeles, Calif.: Morgan Kaufmann, 86–94.
- Cooper, G., and Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks from data. *Machine Learning* 9: 309–347.
- Cooper, G., and Yoo, C. (2000) Causal discovery from a mixture of experimental and observational data, in *Proceedings of the 15th Annual Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufman, San Francisco, CA 116-125.
- Cornfield, J., Haenszel, W., Hammond, E., Lilienfeld, A., Shimkin, M., and Wynder, E. (1959). Smoking and lung cancer: Recent evidence and a discussion of some questions. *Journal of the National Cancer Institute* 22: 173–203.
- Costner, H. (1971). Theory, deduction and rules of correspondence. In *Causal Models in the Social Sciences*, edited by H. Blalock. Chicago: Aldine.
- Costner, H., and Herting, J. (1985). Respecification in multiple indicator models. In *Causal Models in the Social Sciences*, edited by H. Blalock. New York: Aldine, 321–393.
- Costner, H., and Schoenberg, R. (1973). Diagnosing indicator ills in multiple indicator models. In *Structural Equation Models in the Social Sciences*, edited by A. Duncan and O. Goldberger. New York: Seminar Press.
- Cox, D. (1958). *Planning of Experiments*. New York: Wiley.
- Cox, D., and Wermuth, N. (1996). *Multivariate Dependencies: Models, Analysis and Interpretation*. Boca Raton, Fla.: CRC Press.
- Cox, D., and Wermuth, N. (1999). On the generation of the chordless four-cycle, ZUMA-Arbeitsbericht 99/04. *Submitted to Biometrika*.

- Crawford, S. (1994). An application of the Laplace method to finite mixture distributions. *JASA* 89: 259–267.
- Crawford, S., and Fung, R. (1990). An analysis of two probabilistic model induction techniques. *Proceedings of the Third Annual Workshop on Artificial Intelligence and Statistics*, Fort Lauderdale, Fla.
- Dai, H., Korb, K., Wallace C., and Wu, X. (1997). A study of causal discovery with weak links and small samples. *15th International Joint Conference on Artificial Intelligence*, Nagoya, Japan, 23–29 August, 1997.
- Darroch, J., Lauritzen, S., and Speed, T. (1980). Markov fields and log linear interaction models for contingency tables. *Ann. Stat.* 8: 522–539.
- Davis, W. (1988). Probabilistic theories of causation. In *Probability and Causality*, edited by J. Fetzer. Dordrecht: D. Reidel.
- Dawid, A. (1979). Conditional independence in statistical theory (with discussion). *J. Roy. Statist. Soc. Ser. B* 41: 1–31.
- Dawid, A. (1997). Causal inference without counterfactuals. Technical Report, Department of Statistical Science, University College, London.
- de Campos, L., and Huete, J. Approximating causal orderings for Bayesian networks using genetic algorithms and simulated annealing. Technical Report, #DECSAI-990212, Department of Computer Science, University of Grenada, Spain.
- de Fazio, J., Ramsey, J., Roush, T., Gazis, P., and Glymour, C. (2000) Automated Mineral Identification from Reflectance Spectra, Carnegie Mellon University Department of Philosophy Technical Report.
- Dempster, A. (1972). Covariance selection. *Biometrics* 28: 157–175.
- Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc.* 39(Ser. B): 1–38.
- Desjardins, B. (1999). *The Limits of Causal Discovery*. Department of History and Philosophy of Science. Pittsburgh: University of Pittsburgh.
- DiCiccio, T., Kass, R., Raftery, A., and Wasserman, L. (1995). Computing Bayes factors by combining simulation and asymptotic approximations. Pittsburgh: Department of Statistics, Carnegie Mellon University.

- Doll, R., and Hill, A. (1952). A study of the aetiology of carcinoma of the lung. *Brit. Med. Journal* 2: 1271–1286.
- Druzdzel, M., and Glymour, C. (1999). Causal inferences from databases: why universities lose students. *Computation, Causation and Discovery*, edited by C. Glymour and G. Cooper. Cambridge, Mass.: MIT Press.
- Duncan, O. (1975). *Introduction to Structural Equation Models*. New York: Academic Press.
- Duncan, O., Featherman, D., and Duncan, B. (1972). *Socioeconomic Background and Achievement*. New York: Seminar Press.
- Edwards, A. (1976). *An Introduction to Linear Regression and Correlation*. San Francisco: W. H. Freeman.
- Edwards, D. (1995). *Introduction to Graphical Modeling*. New York: Springer-Verlag.
- Edwards, D., and Havranek, T. (1985). A fast procedure for model search in multi-dimensional contingency tables. *Biometrika* 72: 339–351.
- Edwards, D., and Havranek, T. (1987). A fast model selection procedure for large families of models. *J. Amer. Statist. Assoc.* 82: 205–211.
- Edwards, D., and Kreiner, S. (1983). The analysis of contingency tables by graphical models. *Biometrika* 70: 553–565.
- Elby, A. (1992). Should we explain the EPR causally? *Philosophy of Science* 59: 16–25.
- Fienberg, S. (1977). *The Analysis of Cross-classified Categorical Data*. Cambridge, Mass.: MIT Press.
- Fine, A. (1982). Hidden variables, joint probability, and the Bell inequalities. *Physical Review Letters* 48: 291–295.
- Fine, M., Auble, T., Yealy, D., Hanusa, B., Weissfeld, L., Singer, D., Coley, C., Marrie, T., Kapoor, W. (1997). A prediction rule to identify low-risk patients with community-acquired pneumonia. *New England Journal of Medicine* 336: 243–250.
- Fine, T. (1973). *Theories of Probability; an Examination of Foundations*. New York: Academic Press.

- Fisher, F. (1966). *The Identification Problem in Economics*. New York: McGraw-Hill.
- Fisher, F. (1970). A correspondence principle for simultaneous equation models. *Econometrica* 38: 73–92.
- Fisher, R. (1951). *The Design of Experiments*. Edinburgh: Oliver and Boyd.
- Fisher, R. (1959). *Smoking. The Cancer Controversy*. Edinburgh: Oliver and Boyd.
- Flack, V., and Chang, P. (1987). Frequency of selecting noise variables in subset regression analysis: A simulation study. *American Statistician* 41: 84–86.
- Forbes, H., and Tufte, E. (1968). A note of caution in causal modeling. *American Political Science Review* 62: 1258–1264.
- Fox, J. (1984). *Linear Statistical Models and Related Methods, with Applications to Social Research*. New York: Wiley.
- Freedman, D. (1983a). A note on screening regression equations. *American Statistician* 37: 152–155.
- Freedman, D. (1983b). Structural-equation models: A case study. Berkeley: University of California.
- Freedman, D., Navidi, W., and Peters, S. (1986). On the impact of variable selection in fitting regression equations. *On Model Uncertainty and its Statistical Implications. Lecture Notes in Economics and Mathematical Systems*, edited by T. Dijkstra. Berlin: Springer-Verlag, 307.
- Friedman, M. (1957). *A Theory of Consumption Function*. Princeton, N.J.: Princeton University Press.
- Friedman, N. (1997). Learning belief networks in the presence of missing values and hidden variables. *Proceedings of the 14th International Conference on Machine Learning*.
- Friedman, N., and Goldszmidt, M. (1996). Discretization of continuous attributes while learning Bayesian networks. *Proceedings of the 13th International Conference on Machine Learning*. L. Saitta, 157–165.

- Friedman, N., and Goldszmidt, M. (1999a). Data analysis with Bayesian networks: a bootstrap approach. *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*. San Francisco: Morgan Kaufman, 196–205.
- Friedman, N., Goldszmidt, M., and Wyner, A. (1999b). On the application of the bootstrap for computing confidence measures on features of induced Bayesian networks. *Artificial Intelligence and Statistics 99*. San Francisco: Morgan Kaufman, 197–202.
- Friedman, N., Nachman, I., and Pe'er, D. (1999c). Learning Bayesian networks from massive datasets: The “Sparse Candidate Algorithm.” *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*. San Francisco: Morgan Kaufman, 206–215.
- Frydenberg, M. (1990). The chain graph Markov property. *Scandinavian Journal of Statistics* 17: 333–353.
- Fung, R., and Crawford, S. (1990). Constructor: A system for the induction of probabilistic models. *Proceedings of the Eighth National Conference on AI*, Boston: AAAI.
- Furnival, G., and Wilson, R. (1974). Regression by leaps and bounds. *Technometrics* 16: 499–511.
- Galles, D., and Pearl, J. (1995). Testing Identifiability of Causal Effects. *Proceedings of the Eleventh Annual Conference on Uncertainty in Artificial Intelligence*. San Francisco: Morgan Kaufmann, 185–195.
- Galles, D., and Pearl, J. (1998a). Axioms of causal relevance. *Artificial Intelligence* 97: 9–43.
- Galles, D., and Pearl, J. (1998b). An axiomatic characterization of causal counterfactuals. *Foundations of Science* 3: 151–182.
- Geiger, D. (1990). *Graphoids: A Qualitative Framework for Probabilistic Inference*. Los Angeles: University of California.
- Geiger, D., Heckerman, D., and Meek, C. (1996). Asymptotic model selection for directed networks with hidden variables. *Proceedings of the Twelfth Conference on Uncertainty in AI*. San Francisco: Morgan Kaufmann, 283–290.

- Geiger, D., and Heckerman, D. (1991). Advances in probabilistic reasoning. *Proceedings of the Seventh Annual Conference on Uncertainty in AI*. San Francisco: Morgan Kaufmann, 118–126.
- Geiger, D., and Heckerman, D. (1994). Learning Gaussian networks. *Proceedings of the Tenth Conference on Uncertainty in AI*. San Francisco: Morgan Kaufmann, 235–243.
- Geiger, D., and Heckerman, D. (1995). A characterization of the Dirichlet distribution applicable to learning Bayesian networks. Redmond, Wash.: Microsoft Research.
- Geiger, D., and Meek, C. (1999). Quantifier elimination for statistical problems, *Proceedings of the 15th Annual Conference on Artificial Intelligence and Statistics*, San Francisco, Morgan Kaufman, 226–235.
- Geiger, D., and Pearl, J. (1989a). Axioms and Algorithms for Inferences Involving Conditional Independence. Los Angeles: Cognitive Systems Laboratory, University of California.
- Geiger, D., and Pearl, J. (1989a). Logical and Algorithmic Properties of Conditional Independence and Qualitative Independence. Los Angeles, Cognitive Systems Laboratory, University of California.
- Geiger, D., Heckerman, D., King, H., and Meek, C. (1999). On the geometry of DAG models with hidden variables. *Artificial Intelligence and Statistics 99*. San Francisco: Morgan Kauffman.
- Geiger, D., Verma, T., and Pearl, J. (1990). Identifying independence in Bayesian Networks. *Networks* 20: 507–533.
- Geisser, S., and Eddy, W. (1979). A predictive approach to model selection. *JASA* 74: 153–160.
- Gelfand, A., and Smith, A. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association* 85: 398–409.
- Gelman, A., Carlin, J., Stern, H., and Rubin, D. (1995). *Bayesian Data Analysis*. London: Chapman and Hall.
- Gelman, A., Meng, X., and Stern, H.S. (1996). Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statistica Sinica* 6: 733–807.
- Gelman, A., and Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* 7: 457–511.

- Geman, S., and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6: 721–742.
- Geweke, J., Meese, R., and Dent, W. (1983). Comparing alternative tests of causality in temporal systems. *Journal of Econometrics* 21: 161–194.
- Glymour, C. (1983). Social science and social physics. *Behavioral Science* 28: 126–133.
- Glymour, C., Scheines, R., and Spirtes, P. (1989). Why Aviators Leave the Navy: Applications of Artificial Intelligence Procedures in Manpower Research, Naval Personnel Research Development Center.
- Glymour, C., Spirtes P., and Scheines, R. (1991). From probability to causality. *Philosophical Studies* 64: 1–36.
- Glymour, C., Spirtes, P., and Scheines, R. (1991). Independence relations produced by parameter values. *Philosophical Topics* 18: 55–70.
- Glymour, C., Scheines, R., Spirtes, P., and Kelly, K. (1987). *Discovering Causal Structure*. San Diego, Calif.: Academic Press.
- Gold, E. (1965). Limiting recursion. *Journal of Symbolic Logic* 30: 27–48.
- Gold, E. (1967). Language identification in the limit. *Information and Control* 10: 447–474.
- Goldszmidt, M., and Pearl, J. (1992). Rank-based systems: A simple approach to belief revision, belief update, and reasoning about evidence and actions. *Proceedings of the Third International Conference on Knowledge Representation and Reasoning*, edited by B. Nebel, C. Rich, W. Swartout. San Francisco: Morgan Kaufman, 661–672.
- Goldberg, A., and Duncan, O., eds. (1973). *Structural Equation Models in the Social Sciences*. New York: Seminar Press.
- Goodman, L. (1973a). The analysis of multidimensional contingency tables when some variables are posterior to others: A modified path analysis approach. *Biometrika* 60: 179–192.
- Goodman, L. (1973b). Causal analysis of data from panel studies and other kinds of surveys. *American Journal of Sociology* 78: 1135–1191.

- Granger, C. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37: 424–438.
- Greene, T., and Ernhart, C. (1993). Dentine lead and intelligence prior to school entry: A statistical sensitivity analysis. *Journal of Clinical Epidemiology* 46: 323–329.
- Greenland, S. (1989). Modelling variable selection in epidemiologic analysis. *American Journal of Public Health* 79: 340–349.
- Griffiths, W., Hill, R., and Pope, P. (1987). Small sample properties of probit model estimators. *JASA* 82: 929–937.
- Güven D., and Bessler, D. (1997). A note on directed graphs and time series data. Submitted to *Econometric Reviews*.
- Güven, D., and Tolun, M. (1991). Analyzing the inter and intra vocabulary performances in isolated speech recognition. *Journal of American Voice Input/Output Society* 10: 19–37.
- Haberman, S. (1979). *Analysis of Qualitative Data*. San Diego, Calif.: Academic Press.
- Harary, F., Norman R., and Cartwright, D. (1965). *Structural Models: An Introduction to the Theory of Directed Graphs*. New York: Wiley.
- Harary, F., and Palmer, E. (1973). *Graphical Enumeration*. New York: Academic Press.
- Haughton, D. (1988). On the choice of a model to fit data from an exponential family. *Annals of Statistics* 16: 342–355.
- Hausman, D. (1984). Causal priority. *Nous* 18: 261–279.
- Hausman, D. (1998). *Causal Asymmetries*. Cambridge: Cambridge University Press.
- Hausman, D., and Woodward, J. (in press). The causal Markov condition. *British Journal for Philosophy of Science*.
- Havranek, T. (1984). A procedure for model search in multi-dimensional contingency tables. *Biometrics* 40: 95–100.
- Heckerman, D. (1995). A Bayesian approach for learning causal networks. *Proceedings of the Eleventh Conference on Uncertainty in AI*. San Francisco: Morgan Kaufmann, 274–284.

- Heckerman, D. (1998). A tutorial on learning with Bayesian networks. In *Learning in Graphical Models*, edited by M. Jordan. Cambridge, Mass.: MIT Press.
- Heckerman, D., and Shacter, R. (1995). Decision-theoretic foundations for causal reasoning. *Journal of Artificial Intelligence Research* 3: 405–430.
- Heckerman, D., and Geiger, D. (1996). Likelihoods and Priors for Bayesian Networks. Redmond, Wash.: Microsoft Research.
- Heckerman, D., Geiger, D, and Chickering, D. (1994). Learning Bayesian networks: The combination of knowledge and statistical data, *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence*. San Francisco: Morgan Kaufmann, 293–302.
- Heckerman, D., Geiger, D., and Chickering, D. (1995). Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning* 20: 197–243.
- Heckerman, D., Meek, C., and Cooper, G. (1999) A Bayesian approach to causal discovery. *Computation, Causation and Discovery*, edited by C. Glymour and G. Cooper. Cambridge, Mass.: MIT Press.
- Heise, D. (1975). *Causal Analysis*. New York: Wiley.
- Herskovits, E. (1991). Computer-based probabilistic network construction. Medical Information Sciences. Stanford, Calif.: Stanford University.
- Herskovits, E. (1992). Computer Based Probabilistic-Network Construction. Computer Science and Medicine, Stanford University.
- Herskovits, E., and Cooper, G. (1990). Kutato: An entropy-driven system for construction of probabilistic expert systems from databases, *Proceedings of the Sixth Conference on Uncertainty in AI*. Mountain View, Calif.: Association for Uncertainty in AI.
- Herting, J., and Costner, J. (1985). Respecification in multiple indicator models. *Causal Models in the Social Sciences*, edited by H. Blalock. New York: Aldine.
- Hocking, R., and Leslie, R. (1967). Selection of the best subset in regression analysis. *Technometrics* 9: 531–540.
- Hojsgaard, S., and Thiessan, B. (1995). Block recursive models induced from observations and statistical techniques, *Computational Statistics and Data Analysis*. 19: 155-175.

- Holland, P., and Rosenbaum, P. (1986). Conditional association and unidimensionality in monotone latent variable models. *Annals of Statistics* 14: 1523–1543.
- Hoogland, J., and Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods and Research* 26: 329–368.
- Hosmer, D., and Lemeshow, S. (1989). *Applied Logistic Regression*. New York: Wiley.
- Hu, L., Bentler, P., and Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin* 112: 351–362.
- Humphreys, P., and Freedman, D. (1996). The grand leap. *British Journal for the Philosophy of Science* 47: 113–118.
- Iwasaki, Y., and Simon, H. (1994). Causality and model abstraction. *Artificial Intelligence*, 67, 143–194.
- James, L., Mulaik, S., and Brett, J. (1982). *Causal Analysis: Assumptions, Models and Data*. Beverly Hills, Calif.: Sage Publications.
- Jeffreys, H. (1957). *Scientific Inference*. New York: Cambridge University Press.
- Jensen, F., Lauritzen, S., and Oleson, K. (1990). Bayesian updating in recursive graphical models by local computations. *Computational Statistics Quarterly* 4: 269–282.
- Jordan, M. (1998). *Learning in Graphical Models*. NATO Science Series D, Behavioral Social Sciences, vol. 89. Kluwer. Also published in 1999 by MIT Press, Cambridge, Mass.
- Jöreskog, K. (1973). A general method for estimating a linear structural equation. *Structural Equation Models in the Social Sciences*. A. Goldberger and O. Duncan. New York: Seminar Press.
- Jöreskog, K. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika* 43: 443–447.
- Jöreskog, K. (1981). Analysis of covariance structures. *Scandinavian Journal of Statistics* 8: 65–92.
- Jöreskog, K., and Sörbom, D. (1984). *LISREL VI User's Guide*. Mooresville, Ind.: Scientific Software.

- Jöreskog, K., and Sörbom, D. (1990). Model search with TETRAD II and LISREL. *Sociological Methods and Research* 19: 93–106.
- Jöreskog, K. G., and Sörbom, D. (1993). *LISREL 8: User's reference guide*. Chicago: Scientific Software International.
- Kadane, J., and Sedransk, N. (1996). *Bayesian Ethics in a Clinical Trial Design*, edited by J. Kadane. New York: Wiley.
- Kadane, J., and Seidenfeld, T. (1990). Randomization in a Bayesian Perspective. *Journal of Statistical Planning and Inference* 25: 329–345.
- Kano, Y., and Harada, A. (in press). Stepwise variable selection in factor analysis. *Psychometrika*.
- Kass, R., and Raftery, A. (1995). Bayes factors. *JASA* 90: 773–795.
- Kass, R., and Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *JASA* 90: 928–934.
- Kass, R., Tierney, L., and Kadane, J. (1988). Asymptotics in Bayesian computation. *Bayesian Statistics 3*, edited by J. Bernardo, M. DeGroot, D. Lindley, and A. Smith. Oxford: Oxford University Press.
- Katsumo, H., and Mendelson, A. (1991). On the difference between updating a knowledge base and revising it. *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference*, Boston, Mass.
- Kelley, T. (1928). *Crossroads in the Mind of Man*. Stanford: Stanford University Press.
- Kelly, K. (1996). *The Logic of Reliable Inquiry*. New York: Oxford University Press.
- Kendall, M. (1948). *The Advanced Theory of Statistics*. London: Charles Griffin and Co.
- Kenny, D. (1979). *Correlation and Causality*. New York: Wiley.
- Kiiveri, H. (1982). *A Unified Approach to Causal Models*, University of Western Australia.
- Kiiveri, H., Speed, T., and Carlin, J. (1984). Recursive causal models. *Journal of the Australian Mathematical Society* 36: 30–52.

- Kiiveri, H., and Speed, T. (1982). Structural analysis of multivariate data: A review. *Sociological Methodology*, edited by S. Leinhardt. San Francisco: Jossey-Bass.
- Klein, L. (1961). *An Econometric Model of the United Kingdom*. Oxford: Oxford University, Institute of Statistics.
- Kleinbaum, D, Kupper, L., and Morgenstern, H. (1982). *Epidemiologic Research*. Belmont, Calif.: Lifetime Learning Publications.
- Klepper, S., and Leamer, E. (1984). Consistent sets of estimates for regressions with errors in all variables. *Econometrica* 52: 163–183.
- Klepper, S. (1988). Regressor diagnostics for the classical errors-in-variables model. *Journal of Econometrics* 37: 225–250.
- Klir, G., and Parviz, B. (1986). General reconstruction characteristics of probabilistic and possibilistic systems. *International Journal of Man-Machine Studies* 25: 367–397.
- Kohn, M. (1969). *Class and Conformity*. Homewood, Ill.: Dorsey Press.
- Korb, K., and Wallace, C. (1997). In search of the philosopher's stone: remarks on Humphreys and Freedman's critique of causal discovery. *British Journal for the Philosophy of Science* 48: 543–553.
- Koster, J. (1995). Gibbs-factorization and the Markov property. Unpublished manuscript.
- Koster, J. (1996). Markov properties of nonrecursive causal models. *Annals of Statistics* 24: 2148–2178.
- Koster, J. (forthcoming). On the validity of the Markov interpretation of path diagrams of Gaussian structural equations systems with correlated errors. *Scandinavian Journal of Statistics*.
- Kullback, S. (1959, 1968). *Information Theory and Statistics*. New York: Wiley.
- Kullback, S. (1967). A lower bound for discrimination information in terms of variation. *IEEE Trans. Information Theory* 13: 126–127.
- Kullback, S. (1968). Probability densities with given marginal. *Annals of Mathematical Statistics* 39: 79–86.

Lauritzen, S., Speed, T., and Vijayan, K. (1978). Decomposable Graphs and Hypergraphs. University of Copenhagen, Institute of Mathematical Studies. Preprint 9.

Lauritzen, S. (1996). *Graphical Models*. Oxford: Oxford University Press.

Lauritzen, S. (2000). Causal Inference from Graphical Models, in *Complex Stochastic Systems*, edited by O. Barndorff-Nielsen, O. Cox, and C. Klippelberg. London: Chapman and Hall.

Lauritzen, S., and Wermuth, N. (1989). Graphical models for association between variables, some of which are qualitative and some quantitative. *Annals of Statistics* 17: 31–57.

Lauritzen, S., Dawid, A., Larsen, B., and Leimer, H. (1990). Independence properties of directed Markov fields. *Networks* 20: 491–505.

Lawley, D., and Maxwell, A. (1971). *Factor Analysis as a Statistical Method*. London: Butterworth.

Lazarsfeld, P., and Henry, N. (1968). *Latent Structure Analysis*. Boston: Houghton Mifflin.

Leamer, E. (1978). *Specification Searches: Ad Hoc Inference with Non-experimental Data*. New York: Wiley.

Lee, S. (1985). Analysis of covariance and correlation structures. *Computational Statistics and Data Analysis* 2: 279–295.

Lee, S. (1987). Model Equivalence in Covariance Structure Modeling. Department of Psychology, Ohio State University.

Lemmer, J. 1996. The Causal Markov Condition, Fact or Artifact. *SIGART Bulletin* 7: 3–16.

Lewis, D. (1973a). Causation. *Journal of Philosophy* 70: 556–572.

Lewis, D. (1973b). *Counterfactuals*. Cambridge, Mass.: Harvard University Press.

Lilienfeld, A. (1983). The Surgeon General's "Epidemiologic Criteria for Causality." A Criticism of Burch's Critique 36: 837–845.

- Linthurst, R. A. (1979). Aeration, nitrogen, pH and salinity as factors affecting *Spartina Alterniflora* growth and dieback. North Carolina State University.
- Lohmoller, J. (1989). *Latent Variable Path Modeling with Partial Least Squares*. Heidelberg, Physica-Verlag.
- Long, J. (1983a). *Confirmatory Factory Analysis*. Beverly Hills, Calif.: Sage Publications.
- Long, J. (1983b). *Covariance Structure Models*. Beverly Hills, Calif.: Sage Publications.
- Luijben, T., Boomsma, A., and Molenaar, I. (1986). Modification of factor analysis models in covariance structure analysis. A Monte Carlo Study. *On Model Uncertainty and its Statistical Implications*. Berlin: Springer-Verlag. 307.
- MacCallum, R. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin* 100: 107–120.
- Mackie, J. (1974). *The Cement of the Universe*. New York: Oxford University Press.
- Madigan, D. (1999). Bayesian graphical models, intention to treat, and the Rubin causal model. *Artificial Intelligence and Statistics 99*. San Francisco: Morgan Kaufman, 123–132.
- Madigan, D., Garvin, J., and Raftery, A. (1995). Eliciting prior information to enhance the predictive performance of Bayesian graphical models. *Communications in Statistics: Theory and Methods* 24: 2271–2292.
- Madigan, D., Raftery, A., Volinsky, C., and Hoeting, J. (1996). Bayesian model averaging. *AAAI Workshop on Integrating Multiple Learned Models*, Portland, Ore.
- Madigan, D., and York, J. (1995). Bayesian graphical models for discrete data. *International Statistical Review* 63: 215–232.
- Mallows, C. (1973). Some comments on C_p . *Technometrics* 15: 661–676.
- Mani, S., and Cooper, G. (forthcoming) A study in causal discovery from population-based infant birth and death records. *Proceedings of the Annual Fall Symposium of the American Medical Informatics Association*. Hanley and Belfus, Philadelphia.
- Manski, C. (1995). *Identification Problems in the Social Sciences*. Harvard University Press, Cambridge, Mass.

- Mardia, K., Kent, J., and Bibby, J. (1979). *Multivariate Analysis*. New York: Academic Press.
- Maruyama, G., and McGarvey, B. (1980). Evaluating causal models: An application of maximum likelihood analysis of structural equations. *Psychological Bulletin* 87: 502–512.
- MathSoft (1993). *Mathcad 4.0. User's guide. Windows version*. Cambridge, Mass.
- Matu_, F. (1995). Conditional independence properties among four random variables II. *Combinatorics, Probability, and Computing* 4, 407–417.
- Matu_, F., and Studen_, M. (1995). Conditional independence properties among four random variables I. *Combinatorics, Probability, and Computing* 4: 269–278.
- Maudlin, T. (1994). *Quantum Non-Locality and Relativity*. Blackwell, Cambridge, UK.
- McKenna, M., and Shipley, W. 1999. Interacting determinants of interspecific relative growth: empirical patterns and a theoretical explanation. *Ecoscience* (in press).
- McLachlan, G., and Krishnan, T. (1997). *The EM Algorithm and Extensions*. New York: Wiley.
- McPherson, J., Welch, S., and Clark, C. (1977). The stability and reliability of political efficacy: Using path analysis to test alternative models. *American Political Science Review* 71: 509–521.
- Meek, C. (1995). Strong completeness and faithfulness in Bayesian networks. *Proceedings of the Eleventh Conference on Uncertainty in AI*. San Francisco: Morgan Kaufmann, 411–418.
- Meek, C., and Glymour, C. (1994). Conditioning and intervening. *British Journal for Philosophy of Science* 45: 1001–1021.
- Meng, X., and Rubin, D. (1991). Using EM to obtain asymptotic variance-covariance matrices: The SEM algorithm. *JASA* 86: 899–909.
- Miller, J., Slomczynski, K., and Schoenberg, R. (1981). Assessing comparability of measurement in cross-national research: Authoritarian-conservatism in different sociocultural settings. *Social Psychology Quarterly* 44: 178–191.
- Miller, R., Jr. (1981). *Simultaneous Statistical Inference*. New York: McGraw-Hill.

- Miller, W., and Stokes, D. (1963). Constituency influence in Congress. *American Political Science Review* 1963: 45–456.
- Mitchell, T. (1977). Version spaces: A candidate elimination approach to rule learning. *Proceedings of the Fifth International Joint Conference on AI*, Pittsburgh.
- Mitchell, T., and Beauchamp, J. (1988). Bayesian variable selection in linear regression. *JASA* 83: 1023–1032.
- Monti, S., and Cooper, G. (1998). A multivariate discretization method for learning bayesian networks from mixed data. *Proceedings of the Fourteenth Annual Conference on Uncertainty in Artificial Intelligence*. San Francisco: Morgan Kaufmann, 404–413.
- Mosteller, F., and Tukey, J. (1977). *Data Analysis and Regression, A Second Course in Regression*. Reading, Mass.: Addison-Wesley.
- MRFIT Research Group (1982). Multiple risk factor intervention trial; risk factor changes and mortality results. *JAMA* 248: 1465–1477.
- Muthen, B. (1984). A general structural equation model with dichotomous, ordered categorical and continuous latent variable indicators. *Psychometrika* 49: 115–132.
- Neal, R. (1993). Probabilistic inference using Markov chain Monte Carlo methods. Toronto: Department of Computer Science, University of Toronto.
- Neal, R. (2000). On deducing conditional independence from d-separation in causal graphs with feedback: The Uniqueness Condition is not sufficient. *Journal of Artificial intelligence Research*, 12:87-91.
- Neapolitan, R. (1990). *Probabilistic Reasoning in Expert Systems*. New York: Wiley.
- Needleman, H., Geiger, S., and Frank, R. (1985). Lead and IQ Scores: A Reanalysis. *Science* 227: 701–704.
- Neyman, J. (1935). Statistical problems with agricultural experimentation. *J. Roy. Stat. Soc. Suppl.* 2: 107–180.
- Osherson, D., Stob, T., and Weinstein, S. (1986). *Systems That Learn*. Cambridge, Mass.: MIT Press.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. San Mateo, Calif.: Morgan Kaufmann.

- Pearl, J. (1994). A probabilistic calculus of actions. *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence*. Seattle: Morgan Kaufmann, 454–462.
- Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika* 82: 669–709.
- Pearl, J. (1999a). Graphs, structural models and causality. *Computation, Causation and Discovery*. C. Glymour and G. Cooper. Cambridge, Mass.: MIT Press.
- Pearl, J. (1999b). Reasoning with cause and effect. UCLA Cognitive Science Laboratory Technical Report, R–265.
- Pearl, J. (forthcoming). Probabilities of causation: Three counterfactual interpretations and their identification. *Synthese*.
- Pearl, J., and Dechter, R. (1989). *Learning structure from data: A survey*. proceedings COLT '89, Santa Cruz, Calif., 230–423.
- Pearl, J., and Dechter, R. (1996). Identifying independencies in causal graphs with feedback. *Proceedings of the Twelfth Annual Conference on Uncertainty in Artificial Intelligence*. San Francisco: Morgan Kaufman, 420–426.
- Pearl, J., and Robins, J. (1995). Probabilistic evaluation of sequential plans from causal models with hidden variables. *Proceedings of the Eleventh Annual Conference on Uncertainty in Artificial Intelligence*. San Francisco: Morgan Kaufmann, 444–453.
- Pearl, J., and Tarsi, M. (1986). Structuring causal trees. *Journal of Complexity* 2: 60–77.
- Pearl, J., and Verma, T. (1987). The Logic of Representing Dependencies by Directed Graphs. Los Angeles: UCLA Cognitive Systems Laboratory.
- Pearl, J., and Verma, T. (1990). A Formal Theory of Inductive Causation. Los Angeles: UCLA Cognitive Systems Laboratory, Computer Science Department.
- Pearl, J., and Verma, T. (1991). A theory of inferred causation. *Proceedings of the Second International Conference of Representation and Reasoning*, San Francisco: Morgan Kaufmann.
- Pearl, J., Geiger, D., and Verma, T. (1990). The logic of influence diagrams. *Influence Diagrams, Belief Nets and Decision Analysis*, edited by R. Oliver and J. Smith. New York: Wiley.

- Pratt, J., and Schlaifer, R. (1988). On the interpretation and observation of laws. *Journal of Econometrics* 39: 23–52.
- Putnam, H. (1965). Trial and error predicates and a solution to a problem of Mostowski. *Journal of Symbolic Logic* 30: 49–57.
- Pyankov, V., Kondratchuk, A., and Shipley, W. (1999). Leaf structure and specific leaf mass: the alpine desert plants of the Eastern Pamirs (Tadjikistan). *The New Phytologist* (in press).
- Raftery, A. (1993). Bayesian model selection in structural equation models. *Testing Structural Equation Models*, edited by K. A. Bollen and J. S. Long. Newbury Park, Calif.: Sage, 163–180.
- Raftery, A. (1994). Bayesian model selection in social research. Working paper no. 94-12, University of Washington, Center for Studies in Demography and Ecology.
- Raftery, A. (1995). Bayesian model selection in social research. In *Sociological Methodology*, edited by P. Marsden. Cambridge, Mass.: Blackwell.
- Raftery, A. (1996). Hypothesis testing and model selection. *Markov Chain Monte Carlo in Practice*, edited by W. R. Gilks, S. Richardson, and D. Spiegelhalter. London: Chapman and Hall, 163–187.
- Ramoni, M (1996) Learning Bayesian networks from incomplete databases. *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufman, 401–408.
- Rawlings, J. (1988). *Applied Regression Analysis*. Belmont, Calif.: Wadsworth.
- Reichenbach, H. (1956). *The Direction of Time*. Berkeley: University of California Press.
- Reiss, I., Banwart, A., and Forman, H. (1975). Premarital contraceptive usage: A study and some theoretical explorations. *Journal of Marriage and the Family* 37: 619–630.
- Richardson, T. (1996a). Models of Feedback: Interpretation and Discovery. Department of Philosophy. Pittsburgh: Carnegie Mellon University.
- Richardson, T. (1996b) A discovery algorithm for directed cyclic graphs. *Proceedings of the 12th Conference of Uncertainty in AI*, Portland, Ore.: Morgan Kaufmann, 454–461.

- Richardson, T. (1996c). A polynomial-time algorithm for deciding Markov equivalence of directed cyclic graphical models. *Proceedings of the 12th Conference of Uncertainty in AI*, Portland, Ore.: Morgan Kaufmann, 462–469.
- Richardson, T. (1998). Chain graphs and symmetric associations. *Learning in Graphical Models*. M. Jordan. Cambridge, Mass.: MIT Press.
- Richardson, T. (1999b). A local Markov property for acyclic directed mixed graphs. *Proceedings of 52nd Session of the International Statistical Institute*, Aug. 10–18, Helsinki, Finland.
- Richardson, T., Bailer, H., and Banerjee, M. (1999). Specification searches using MAG models. *Proceedings of 52nd Session of the International Statistical Institute*, Aug. 10–18, Helsinki, Finland.
- Richardson, T., and Spirtes, P. (1999). Parameterizing and Scoring Mixed Ancestral Graphs. Carnegie Mellon Department of Philosophy Technical Report Phil 102.
- Rindfuss, R., Bumpass, L., and St. John, C. (1980). Education and fertility: Implications for the roles women occupy. *American Sociological Review* 45: 431–447.
- Rissanen, J. (1987). Stochastic complexity (with discussion). *J. Roy. Stat. Soc.* 49(Ser. B): 223–239 and 253–265.
- Robins, J. (1986). A new approach to causal inference in mortality studies with sustained exposure period—applications to control of healthy workers survivor effect. *Mathematical Modelling* 7: 1393–1512.
- Robins, J. (1987). Addendum to “A new approach to causal inference in mortality studies with sustained exposure period – applications to control of healthy workers survivor effect.” *Computers and Mathematics with Applications* 14: 923–945
- Robins, J. (1993). Analytic methods for estimating HIV-treatment and cofactor effects. *Methodological Issues in AIDS Mental Health Research*, edited by D. Ostrow and R. Kessler. New York: Plenum, 213–290.
- Robins, J. (1994). Correcting for noncompliance in randomized trials using structural nested mean models. *Communications in Statistics* 23: 2379–2412.
- Robins, J. (1995). Discussion of “Causal Diagrams for Empirical Research” by J. Pearl. *Biometrika* 82: 695–698.

- Robins, J. (1997). Causal inference from complex longitudinal data. *Latent Variable Modeling and Applications to Causality*, M. Berkane. Lecture Notes in Statistics 120. New York: Springer Verlag.
- Robins, J. (1998). Marginal Structural models. *Proceedings of the American Statistical Association*, Chicago, 1–10.
- Robins, J. (1999). Testing and estimation of direct effects by reparameterizing directed acyclic graphs with structural nested models. *Computation, Causation and Discovery*. C. Glymour and G. Cooper. Cambridge, Mass.: MIT Press.
- Robins, J., and Greenland, S. (1989). The probability of causation under a stochastic model for individual risk. *Biometrics* 45: 1125–1138.
- Robins, J., Scheines, R., Spirtes, P., and Wasserman, L. (1999). The Limits of Causal Knowledge. Carnegie Mellon University Philosophy Department Technical Report Phil 97.
- Robins, J., and Wasserman, L. (1999). On the impossibility of inferring causation from association without background knowledge. In *Computation, Causation, and Discovery*, edited by C. Glymour and G. Cooper. Cambridge, Mass.: MIT Press.
- Rodgers, R., and Maranto, C. (1989). Causal models of publishing productivity in psychology. *Journal of Applied Psychology* 74: 636–649.
- Rose, G. Hamilton, P. Colwell, L., and Shipley, J. (1982). A randomised controlled trial of anti-smoking advice: 10-year results. *Journal of Epidemiology and Community Health* 36: 102–108.
- Rosenbaum, P. (1984). From association to causation in observational studies. *JASA* 79: 41–48.
- Rosenbaum, P. (1995). *Observational Studies*. New York: Springer-Verlag.
- Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66: 688–701.
- Rubin, D. (1977). Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics* 2: 1–26.
- Rubin, D. (1978). Bayesian inference for causal effects: The role of randomizations. *Annals of Statistics* 6: 34–58.

- Rubin, D. (1986). Comment: Which ifs have causal answers. *Journal of the American Statistical Association* 81: 396.
- Russell, S., Binder, J., Koller, D., and Kanazawa, K. (1995). Local learning in probabilistic networks with hidden variables. *Proceedings of the International Joint Conference on AI*. San Francisco: Morgan Kaufmann.
- Salmon, W. (1980). Probabilistic causality. *Pacific Philosophical Quarterly* 61: 50–74.
- Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton, N.J.: Princeton University Press.
- Saris, W., and Stronkhorst, H. (1984). *Causal Modeling in Nonexperimental Research*. Amsterdam, Sociometric Research Foundation.
- Scheines, R. (1988). Automating creativity. *Aspects of Artificial Intelligence*. J. Fetzer. Boston: Kluwer.
- Scheines, R., Spirtes, P., Glymour, G., and Sorensen, S. (1990). Causes of Success and Satisfaction Among Naval Recruiters. San Diego, Calif.: Navy Personnel Research Development Center.
- Scheines, R., Spirtes, P., and Glymour, C. (1990). A qualitative approach to causal modeling. *Qualitative Simulation Modeling and Analysis*. P. Luker and P. Fishwick. New York: Springer-Verlag, 72–97.
- Scheines, R. (1993). Unidimensional linear latent variable models. *Technical Report CMU-PHIL-39*. Carnegie Mellon University, Pittsburgh.
- Scheines, R., Spirtes, P., Glymour, C., and Meek, C. (1994). *Tetrad II: User's Manual*. Hillsdale, N.J.: Lawrence Erlbaum.
- Scheines, R. (1997). Estimating latent causal influence: TETRAD II model selection and Bayesian parameter estimation. *Proceedings of the 6th International Workshop on Artificial Intelligence and Statistics*. Fort Lauderdale, Fla., 445–456.
- Scheines, R., Boomsma, A., and Hoijtink, H. (1997). The multimodality of the likelihood function in structural equation models. Pittsburgh.: Department of Philosophy, Carnegie Mellon University.

- Scheines, R., Spirtes, P., Glymour, C., Meek, C., and Richardson, T. (1998). The TETRAD Project: constraint based aids to causal model specification. *Multivariate Behavioral Research* 33: 65–118.
- Scheines, R., Boomsma, A., and Hoijtink, H. (1999). Bayesian Estimation and Testing of Structural Equation Models. *Psychometrika* 64: 37–52.
- Scheines, R., and Spirtes, P. (1992). Finding latent variable models in large data bases. *International Journal of Intelligent Systems* 7: 609–622.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* 6: 461–464.
- Sclove, S. (undated). On Criteria for Choosing a Regression Equation for Prediction. Pittsburgh: Department of Statistics, Carnegie Mellon University.
- Settimi, R., and Smith, J. Geometry, moments, and Bayesian networks with hidden variables. *Artificial Intelligence and Statistics* 99. San Francisco: Morgan Kaufman, 293–298.
- Sewell, W., and Shah, V. (1968). Social class, parental encouragement, and educational aspirations. *American Journal of Sociology* 73: 559–572.
- Shafer, S. (1996). *The Art of Causal Conjecture*. Cambridge, Mass.: MIT Press.
- Shigemasa, K. (1999). Computing Bayes factors for structural equation models. *Proceedings of the 1999 Meetings of the European Psychometric Society*, Luneburg, Germany.
- Shipley, W. (1995). Structured interspecific determinants of specific leaf area in 34 species of herbaceous angiosperms. *Functional Ecology* 9: 312–319.
- Shipley, W. (1997). Exploratory path analysis with applications in ecology and evolution. *American Naturalist* 149: 1113–1138.
- Shipley, W. (1999). Exploring hypothesis space: examples from organismal biology. *Computation, Causation and Discovery*, edited by C. Glymour and G. Cooper. Cambridge, Mass.: MIT Press.
- Shipley, W. (1999). Testing causal explanations in organismal biology: causation, correlation and structural equation modelling. *Oikos* (in press).

- Simon, H. (1953). Causal ordering and identifiability. *Studies in Econometric Methods*. Hood and Koopmans. New York: Wiley, 49–74.
- Simon, H. (1954). Spurious correlation: a causal interpretation. *JASA* 49: 467–479.
- Simon, H. (1977). *Models of Discovery*. Dordrecht: D. Reidel.
- Simpson, C. (1951). The interpretation of interaction in contingency tables. *J. Roy. Statist. Soc.* 13(Ser. B): 238–241.
- Sims, C. (1972). Money, income, and causality. *American Economic Review* 62: 540–552.
- Singh, M., and Valorta, M. (1993). An algorithm for the construction of Bayesian network structure from data. *Proceedings of the Ninth Conference on Uncertainty in AI*. San Francisco: Morgan Kaufmann, 259–265.
- Skyrms, B. (1980). *Causal Necessity: A Pragmatic Investigation of the Necessity of Laws*. New Haven: Yale University Press.
- Smith, A., and Roberts, G. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society* 55(Ser. B): 3–23.
- Sober, E. (1987). The principle of the common cause. *Probability and Causality*. J. Fetzer. Dordrecht: D. Reidel.
- Sorbom, D. (1975). Detection of correlated errors in longitudinal data. *British Journal of Mathematical and Statistical Psychology* 28: 138–151.
- Sosa, E., and Tooley, M., eds. (1993). *Causation*. New York: Oxford University Press.
- Spearman, C. (1904). General intelligence objectively determined and measured. *American Journal of Psychology* 15: 201–293.
- Spence, M. (1973). Job market signalling. *Quarterly Journal of Economics* 87: 355–379.
- Spiegelhalter, D. (1986). Probabilistic reasoning in predictive expert systems. *Uncertainty in Artificial Intelligence*, edited by K. Lemmer and J. Kanal. Amsterdam: North-Holland.

Spiegelhalter, D., and Knell-Jones, R. (1984). Statistical and knowledge-based approaches to clinical decision-support systems. *J. Royal Statist. Soc.* 147(Ser. A): 35–77.

Spiegelhalter, D., and Lauritzen, S. (1990). Sequential updating of conditional probabilities on directed graphical structures. *Networks* 20: 579–605.

Spiegelhalter, D., and Lauritzen, S. (1995). Sequential updating of conditional probabilities on directed graphical structures. *Networks* 20: 579–605.

Spirtes, P. (1989). Fast Geometrical Calculations of Overidentifying Constraints. Pittsburgh: Laboratory for Computational Linguistics, Carnegie Mellon University.

Spirtes, P. (1989). A Necessary and Sufficient Condition for Conditional Independencies to Imply a Vanishing Tetrad Difference. Pittsburgh: Laboratory for Computational Linguistics, Carnegie Mellon University.

Spirtes, P. (1994a). Building causal graphs from statistical data in the presence of latent variables. *Logic, Methodology, and Philosophy of Science IX*, edited by D. Prawitz, B. Skyrms, and D. Westerståhl. Amsterdam: north Holland, 813–832.

Spirtes, P. (1994b). Conditional Independence in Directed Cyclic Graphical Models for Feedback. Pittsburgh: Dept. of Philosophy, Carnegie Mellon University.

Spirtes, P. (1995). Directed cyclic graphical representation of feedback models. *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, San Francisco: Morgan Kaufmann, 491–498.

Spirtes, P. (1996). Discovering Causal Relations Among Latent Variables in Directed Acyclic Graphical Models. Pittsburgh: Department of Philosophy, Carnegie Mellon University.

Spirtes, P., and Cooper, G. (1998). An experiment in causal discovery. *Artificial Intelligence and Statistics '99*, edited by D. Heckerman and J. Whittaker. San Francisco: Morgan Kaufman, 162–168.

Spirtes, P., and Glymour, C. (1988). Latent variables, causal models and overidentifying constraints. *Journal of Econometrics* 39: 175–198.

Spirtes, P., and Glymour, C. (1990). Causal Structure Among Measured Variables Preserved with Unmeasured Variables. Pittsburgh: Laboratory for Computational Linguistics, Carnegie Mellon University.

Spirtes, P., and Meek, C. (1995). Learning Bayesian networks with discrete variables from data. *Proceedings of the First International Conference on Knowledge, Discovery and Data Mining*. San Francisco: Morgan Kaufmann.

Spirtes, P., and Richardson, T. (1996). A polynomial time algorithm for determining DAG equivalence in the presence of latent variables and selection bias. *Proceedings of the 6th International Workshop on Artificial Intelligence and Statistics*, Fort Lauderdale, Fla., 489–500.

Spirtes, P., and Verma, T. (1992). Equivalence of causal models with latent variables. Pittsburgh: Carnegie Mellon University.

Spirtes, P., Glymour, C., and Scheines, R. (1990a). Causality from probability. *Evolving Knowledge in Natural Science and Artificial Intelligence*, edited by J. Tiles et al. London: Pitman, 181–199.

Spirtes, P., Glymour, C., and Scheines, R. (1990b). Causality from probability. *Proceedings of the Conference on Advanced Computing for the Social Sciences*, Williamsburg, Va.

Spirtes P., Glymour C., and Scheines, R. (1991a). An algorithm for fast recovery of sparse causal graphs. *Social Science Computer Review* 9: 62–72.

Spirtes P., Glymour C., and Scheines, R. (1991b). From probability to causality. *Philosophical Studies* 64: 1–36.

Spirtes, P., Glymour, C., and Scheines, R. (1997). Reply to Humphreys' and Freedman's review of Causation, Prediction, and Search. *British Journal for the Philosophy of Science*, 48, 555–568.

Spirtes, P., Glymour, C., Scheines, R., and Sorensen, S. (1990). TETRAD Studies of Data for Naval Air Traffic Controller Trainees. San Diego, Calif.: Navy Personnel Research Development Center.

Spirtes, P., Meek, C., and Richardson, T. (1995). Causal inference in the presence of latent variables and selection bias. *Proceedings of the Eleventh Conference on Uncertainty in AI*, San Francisco: Morgan Kaufmann, 499–506.

Spirtes, P., Meek, C., and Richardson, T. (1999). An algorithm for causal inference in the presence of latent variables and selection bias. *Computation, Causation and Discovery*, edited by C. Glymour and G. Cooper. Cambridge, Mass.: MIT Press.

Spirtes, P., Richardson, T., and Meek, C. (1996). Heuristic greedy search algorithms for latent variable models. *Proceedings of the 6th International Workshop on Artificial Intelligence and Statistics*, Fort Lauderdale, Fla., 481–488.

Spirtes, P., Richardson, T., and Meek, C. (1997). The Dimensionality of Mixed Ancestral Graphs. Technical Report CMU-PHIL–83, Department of Philosophy, Carnegie Mellon University, November, 1997.

Spirtes, P., Richardson, T., Meek, C., Scheines, R., and Glymour, C. (1996). Using D-separation to calculate zero partial correlations in linear models with correlated errors. Technical Report Phil–72, Department of Philosophy, Carnegie Mellon University.

Spirtes, P., Richardson, T., Meek, C., Scheines, R., and Glymour, C. (1998). Using path diagrams as a structural equation modeling tool. *Sociological Methods and Research*, 27, 148–181.

Spirtes, P., Scheines, R., and Glymour, C. (1990). Simulation studies of the reliability of computer aided specification using the TETRAD II, EQS, and LISREL Programs. *Sociological Methods and Research* 19: 3–66.

Spohn, W. (1983). Deterministic and probabilistic reasons and causes. *Methodology, Epistemology, and Philosophy of Science: Essays in Honour of Wolfgang Stegmüller on the Occasion of his 60th Birthday*, edited by H. Putnam, C. Hempel, and W. Essler. Dordrecht: D. Reidel, 371–396.

Spohn, W. (1990). Direct and indirect causes. *Topoi* 9: 125–145.

Spohn, W. (1991). On Reichenbach's principle of the common cause. *Proceedings of the First Pittsburgh-Konstanz Colloquium*.

Spohn, W. (1992). Causal laws are objectifications of inductive schemes. *Theory of Probability*. J. Dubucs. Dordrecht: Kluwer.

Stein, C. (1960). Multiple regression. *Contributions to Probability and Statistics. Essays in Honor of Harold Hotelling*. I. Olkin. Stanford, Stanford University Press.

Stetzl, I. (1986). Changing causal relationships without changing the fit: Some rules for generating equivalent LISREL models. *Multivariate Behavior Research* 21: 309–331.

Strotz, R., and Wold, H. (1960). Recursive versus nonrecursive systems: an attempt at synthesis. *Econometrica* 28: 417–427.

- Studen_, M. (1992). Conditional independence relations have no finite complete characterization. *Information Theory, Statistical Decision Functions and Random Processes: Proceedings of the 11th Prague Conference*, ed. by S. Kubik and J. Vi_ek., Dordrecht: Kluwer, 377–396.
- Studen_, M., and Bouckaert, R. (1998). On chain graph models for descriptions of conditional independence structures. *Annals of Statistics* 26: 1434–1495.
- Suppes, P. (1970). *A Probabilistic Theory of Causality*. Amsterdam, North-Holland.
- Suppes, P., and Zanotti, M. (1981). When are probabilistic explanations possible. *Synthese* 48: 191–199.
- Surgeon General of the United States (1964). *Smoking and Health*. Washington, D.C., U.S. Government Printing Office.
- Surgeon General of the United States (1979). *Smoking and Health*. Washington, D.C., U.S. Government Printing Office.
- Swamy, P. (1971). *Statistical Inference in Random Coefficient Regression Models*. Berlin, Springer-Verlag.
- Tanner, M.A. (1993). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. New York: Springer.
- Thiesson, B. (1995). Score and information for recursive exponential models with incomplete data. Aalborg, Denmark, Institute of Electronic Systems, Aalborg University.
- Thomson, G. (1916). A hierarchy without a general factor. *British Journal of Psychology* 8: 271–281.
- Thomson, G. (1935). On complete families of correlation coefficients and their tendency to zero tetrad-differences: Including a statement of the sampling theory of abilities. *British Journal of Psychology* 26: 63–92.
- Thurstone, L. (1935). *The Vectors of Mind*. Chicago: University of Chicago Press.
- Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics* 22: 1701–1762.

- Timberlake, M., and Williams, K. (1984). Dependence, political exclusion, and government repression: Some cross-national evidence. *American Sociological Review* 49: 141–146.
- Tolun, M., and Guven, D. (1991). Analyzing the inter and intra vocabulary performances in isolated speech recognition. *Journal of American Voice Input/Output Society* 10: 19–37.
- Verma, T. (1987). Causal networks: semantics and expressiveness. Los Angeles, Cognitive Systems Laboratory, University of California.
- Verma, T., and Pearl, J. (1990). Equivalence and synthesis of causal models. *Proceedings of the Sixth Conference on Uncertainty in AI*, Mountain View, Calif.: Association for Uncertainty in AI.
- Verma, T., and Pearl, J. (1990). On equivalence of causal models. Los Angeles, Cognitive Systems Laboratory, University of California.
- Verma, T., and Pearl, J. (1991). Equivalence and synthesis of causal models. Los Angeles, Cognitive Systems Laboratory, University of California.
- Verma, T., and Pearl J. (1990). Causal networks: semantics and expressiveness. *Uncertainty in Artificial Intelligence 4*. T. Levitt R. Shacter, L. Kanal, J. Lemer. North-Holland, Elsevier Science.
- Waldemark, J., and Norqvist, P. In-flight calibration of satellite ion composition data using artificial intelligence methods. *Computation, Causation and Discovery*, edited by C. Glymour and G. Cooper. Cambridge, Mass.: MIT Press.
- Wallace, C., and Freeman, P. (1987). Estimation and inference by compact coding. *J. Roy. Statist. Soc.* 49(Ser. B): 240–265.
- Wedelin D. (1996). Efficient estimation and model selection in large graphical models. *Statistics and Computing* 6: 319–323.
- Weisberg, S. (1985). *Applied Linear Regression*. New York: Wiley.
- Wermuth, N. (1976). Model search among multiplicative models. *Biometrika* 32: 253–363.
- Wermuth, N. (1980). Linear recursive equations, covariance selection and path analysis. *JASA* 75: 963–972.

Wermuth, N., and Lauritzen, S. (1983). Graphical and recursive models for contingency tables. *Biometrika* 72: 537–552.

Wermuth, N., and Lauritzen, S. (1990). On substantive research hypotheses, conditional independence graphs and graphical chain models. *J. Roy. Statist. Soc.* 52(Ser. B): 21–50.

Wermuth, N., Cox, D., and Pearl, J. (1994, revised 1998) Explanations for multivariate structures derived from univariate recursive regressions. Center of Survey Research and Methodology, ZUMA, Mannheim, FRG.

Wermuth, N., Cox, D., Richardson, T., and Glonek, G. (1999). On transforming and generating cyclic graph models. ZUMA Tech report, July 1999.

Wheaton, B., Muthen, B., Alwin, D., and Summers, G. (1977). Assessing Reliability and Stability in Panel Models. *Sociological Methodology*. D. Heise. San Francisco, Jossey-Bass.

Whitney, H. (1957). Elementary structures of real algebraic varieties. *Annals of Mathematics* 66: 545–556.

Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*. New York: Wiley.

Winkler, R. (1967). The assessment of prior distribution in Bayesian analysis. *American Statistical Association Journal* 62: 776–800.

Wise, D. (1975). Academic achievement and job performance. *American Economic Review* 65: 350–366.

Wishart, J. (1928). Sampling errors in the theory of two factors. *British Journal of Psychology*, 19: 180–187.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics* 5: 161–215.

Younger, M. (1978). *Handbook for Linear Regression*. North Scituate, Mass.: Duxbury Press.

Yule, G. (1903). Notes on the theory of association of attributes in statistics. *Biometrika* 2: 121–134.

