

**Discovering Causal Structure:
Artificial Intelligence, Philosophy of Science,
and Statistical Modeling**

by

**Clark Glymour, Richard Scheines, Peter Spirtes,
Kevin Kelly**

August 1986

Report CMU-PHIL-1



**Philosophy
Methodology
Logic**

Pittsburgh, Pennsylvania 15213-3890

DISCOVERING CAUSAL STRUCTURE

Artificial Intelligence, Philosophy of Science and Statistical Modelling

Clark Glymour, Richard Scheines, Peter Spirtes and Kevin Kelly

Department of Philosophy
Carnegie-Mellon University
Pittsburgh, PA. 15213

Copyright (C) 1986 Causal Research Inc.
All Rights Reserved

Page 1 of 1
Date: 10/10/2023
Time: 10:10:10
IP: 192.168.1.1

Table of Contents

1. THE PROBLEMS OF SCIENCE WITHOUT EXPERIMENTS	7
1.1. The Limits of Experimentation	7
1.2. The Limits of Human Judgement	8
1.2.1. The Psychological Problem	9
1.2.2. The Combinatorial Problem	10
1.2.3. Some Examples	13
1.3. The Artificial Intelligence Solution	15
2. THE CASE AGAINST CAUSAL MODELING	21
2.1. The Critical Reaction	21
2.2. Making Sense of Causality	22
2.2.1. Interpreting Causal Claims	24
2.2.2. Baumrind and the Concept of Causality	26
2.3. Causes, Indicators and the Interpretation of Latent Variables	28
2.3.1. General Objections	28
2.3.2. Interpreting Latent Variables	30
2.4. The Importance of Experiment	31
2.4.1. Zeisel on Experimentation and the TARP Experiment	32
2.5. Justifying Assumptions	36
2.5.1. Ling on Causal Modeling	36
2.6. Linear Theories are Literally False	37
2.6.1. The Aims of Linear Models	38
2.6.2. Blau and Duncan's Model	39
2.7. Conclusion	46
3. OBJECTIONS TO DISCOVERY BY COMPUTER	47
3.1. Introduction	47
3.2. The General Objections	48
3.2.1. People Know more than Computers Do	48
3.2.2. The Bayesian Objection	49
3.2.3. The "Fire" Objection	50
3.3. No Peeking	52
3.3.1. Preliminary Remarks	52
3.3.2. Exploratory and Confirmatory Procedures	53
3.3.3. Equivocating over the Model	54
3.3.4. Difficulties with Frequency Interpretations	56
3.3.5. The Worst Case Objection	61
3.3.6. Explanation and Prediction	64
4. CAUSAL AND STATISTICAL MODELS	69
4.1. Introduction	69
4.2. Some Terminology	70
4.3. Directed Graphs and Causal Models	70
4.3.1. Directed Graphs	70
4.3.2. Paths and Treks	71
4.4. Statistical Models from Causal Models	75
4.5. Treks and Coordinating Path Effects	79
4.6. Constraints on Correlations	81
4.6.1. Overidentifying Constraints and Directed Graphs	81
4.6.2. Vanishing Partial Correlations	82

4.6.3. Tetrad Equations	85
4.6.4. Tetrad Equations Derived from Vanishing Partial Correlations	88
4.6.5. Graphs and Constraints	90
4.6.6. Other Constraints	90
4.6.7. Latent Variables and Overidentifying Constraints	91
4.7. Correlated Errors are not Equivalent to Direct Effects	92
4.8. Statistical Issues, Briefly Considered	96
5. THE STRUCTURE AND METHOD OF TETRAD	99
5.1. Introduction	99
5.2. The Methodological Principles That Underlie TETRAD	99
5.2.1. The Inference Problem	99
5.2.2. Implying Constraints That Hold in the Data	100
5.2.3. Not Implying Constraints That Do Not Hold in the Data	100
5.2.4. The Simplicity Principle	101
5.2.5. The Principles Conflict	101
5.3. How the Methodological Principles Are Realized in TETRAD	101
5.3.1. The Constraints That TETRAD Calculates	101
5.3.2. Comparing Models	102
5.3.3. Comparing Elaborations of a Model	102
5.4. A Search Strategy for Finding Good Causal Models	103
5.5. A Sketch of the TETRAD Program	104
5.5.1. TETRAD's Output	104
5.5.2. Tetrad Residuals	106
5.5.3. Partial Correlation Residuals	107
5.5.4. The Statistical Procedures	108
5.5.5. Determining the Constraints Implied by a Model	110
5.5.6. Indices of Fit	110
5.5.7. TETRAD's Automatic Search Procedure	112
5.5.8. An Example	114
5.6. The Value of p Values	127
5.7. TETRAD and Other Search Procedures	128
6. WHAT TETRAD CAN DO	131
6.1. Introduction	131
6.2. Alienation	131
6.3. A Problem Using Simulated Data	134
6.3.1. Study 1	134
6.4. Causal Order from Correlations	135
6.4.1. Monte Carlo Data: I	135
6.4.2. Monte Carlo Data: II	138
6.4.3. Kohn's Study and Temporal Order Among Interview Questions	140
7. USING TETRAD TO DETERMINE CAUSAL ORDER	143
7.1. Introduction	143
7.2. Distinguishing Correlation from Causation	143
7.3. Finding Which Pair of Variables Has an Extra Source of Correlation	147
8. CASE STUDIES	151
8.1. Introduction	151
8.2. Industrial and Political Development	152
8.2.1. Introduction	152
8.2.2. The TETRAD Analysis	153
8.2.3. Goodness of Fit Results	164

8.2.4. Conclusion	168
8.3. Measuring the Authoritarian Personality	170
8.3.1. Introduction	170
8.3.2. The TETRAD Analysis	173
8.3.3. Goodness of Fit Results	183
8.3.4. Predicting Temporal Order	184
8.3.5. Conclusion	186
8.4. Alternatives to Regression Models	187
8.4.1. Introduction	187
8.4.2. The Dependency Theory Example	188
8.4.3. The TETRAD Analysis	190
8.5. Introducing Latent Variables: Longitudinal Data with SAT Scores	193
8.5.1. An Example	193
8.6. The Causal Relations between Constituent Attitudes and Congressional Roll Call Voting	197
8.6.1. Introduction	197
8.6.2. Miller and Stokes' Article	198
8.6.3. Cnudde and McCrone	200
8.6.4. Forbes and Tufte	202
8.6.5. The TETRAD Analysis	204
8.6.6. Goodness of Fit Results	208
8.6.7. Conclusion	209
8.7. The Effects of Summer Head Start	210
8.7.1. Introduction	210
8.7.2. Difficulties with the Magidson-Sorbom Model	213
8.7.3. The TETRAD Analysis	214
8.7.4. Goodness of Fit Results	221
8.7.5. Conclusion	223
8.8. Achievement, Ability, and Approval in School Age Children	224
8.8.1. Introduction	224
8.8.2. The Maruyama and McGarvey Study	224
8.8.3. A TETRAD Search	227
8.9. The Stability of Alienation	239
8.9.1. Introduction	239
8.9.2. The TETRAD Analysis	240
8.9.3. Searching with the Rtrr Chart	244
8.9.4. Goodness of Fit Results	251
8.9.5. Conclusion	253
8.10. A Simulated Case	254
8.10.1. Introduction	254
8.10.2. The TETRAD Analysis	255
8.10.3. Goodness of Fit Results	258
8.10.4. Conclusion	259
9. A BRIEF HISTORY OF HEURISTIC SEARCH IN APPLIED STATISTICS	261
10. MATHEMATICAL FOUNDATIONS	275
10.1. The Algorithm	275
10.1.1. Computing the Implied Equations	275
10.1.2. Suggesting Modifications	277
10.1.2.1. Computing the Rtrr Values	284
10.2. Proofs of Correctness of the Algorithms Employed by TETRAD	285
10.2.1. Introduction	285

10.2.2. Tetrad Equations	285
10.2.3. Proof of Correctness of the Algorithm for Computing Vanishing Partial Correlations	314
11. USING TETRAD WITH EQS AND LISREL	327
11.1. Introduction	327
11.2. LISREL and Its Restrictions	327
11.3. Overcoming the Restrictions	330
11.3.1. An Example: Authoritarian-Conservatism Again	335
11.4. EQS	338
12. Running TETRAD	341
12.1. Introduction	341
12.2. Installing TETRAD	341
12.3. Entering and Exiting TETRAD	341
12.4. Getting Help	342
12.5. Input Files	342
12.5.1. Variable Labels	342
12.5.2. Covariance Files	343
12.5.2.1. Covariances on Each Line	343
12.5.2.2. A Matrix Representation of Covariances	346
12.5.3. Graph Files	347
12.6. Output Files	348
12.7. View and Edit	348
12.8. The Run Command and Menus	348
12.9. User Interrupts	352
12.10. Errors	352
12.10.1. The Form of Error Messages	352
12.10.2. Actions	353
12.10.3. Agreement between Graph and Covariance Files	353
12.10.4. Maximum Number of Variables and Edges	354
12.10.5. Memory	354
12.10.6. Semantic Errors TETRAD Will Not Recognize	355
12.11. TETRAD Commands	355
12.11.1. Types of Commands	356
12.11.2. Switches	356
12.11.3. Prompting	357
12.12. Running TETRAD in Batch Mode	357
12.13. Command Summary	358
12.13.1. Introduction	358
12.13.2. List of Commands	358
12.13.3. Command Summary	359
Appendix I. The Grammar of the Input	377
Index	389

ABOUT THIS BOOK

This book is about a computer program, TETRAD. The program is designed to assist in the search for causal explanations of statistical data. There are often many millions of possible alternative causal explanations for non-experimental data; human researchers can only consider and evaluate a handful of these alternatives, and hope that the best explanation they find is the best explanation there is. That hope is often vain. The TETRAD program uses artificial intelligence techniques to help the investigator to perform a systematic search for alternative causal models, using whatever relevant knowledge may be available.

A version of the TETRAD program, together with several data sets discussed in the book, are included on the floppy disc accompanying this volume. The version of the TETRAD program on the floppy will run on IBM personal computers and on IBM compatible machines. We recommend 512K of RAM, but simple models can be run on machines with less memory. The TETRAD program is intended to be used in conjunction with programs for statistical estimation and testing, such as the LISREL and EQS programs.

The program on the floppy is limited to models with no more than nine variables. A version of the IBM program suitable for larger models (up to 23 variables) is available from Academic Press. Still larger models can be treated using a version of the TETRAD program for the DEC Microvax II Workstation, also available from Academic Press.

Directions for running the program can be found in Chapter 12. Effective use of the program requires study of Chapter 5 and of some of the examples in Chapter 8.

Because the book was written as the program developed, output described in the book's examples may differ in minor ways from output obtained with the version of the program on the floppy disc.

ACKNOWLEDGEMENTS

The development of the TETRAD program has taken the better part of four years. We have been aided during that period both by people and by institutions. The National Science Foundation Program in History and Philosophy of Science provided support which led to the initial formulation of the TETRAD project. The NSF economics program in Measurement Methods helped to fund the development of the computer program. The University of Pittsburgh and Carnegie-Mellon University provided computational resources. A second grant from the History and Philosophy of Science Program at NSF helped to support the writing of this book.

Jill Larken and Preston Covey made equipment available for us, and Robert Daley kindly helped us with some computational needs. Rick Statman and Richmond Thomason gave us some valuable mathematical help. Alison Kost assisted us with the many figures the book contains. Many others have helped us through criticism, encouragement, data or references. We thank William Bielby, Donald Campbell, Herbert Costner, Patrick Doreian, Thomas Ferarro, Steven Fienberg, Daniel Hausman, Max Henrion, Paul Humphreys, Jay Kadane, Melvin Kohn, Patrick Larkey, Peter Machamer, Paul Meehl, Jay Magidson, Merrilee Salmon, Wesley Salmon, Alan Sampson, Kenneth Schaffner, Ronald Schoenberg, Herbert Simon and Jan Zytkow. Teddy Seidenfeld read through much of the manuscript and gave us valuable suggestions. Alison Kost, Martha Harty Scheines, and Jessica Feldman have patiently endured our obsession for several years.

Editors and administrators deserve more credit than they get. This program and book could not have been developed in most places; only the special environment of Carnegie-Mellon University has made it possible, and John Modell and Michael Salomone, respectively Acting Dean and Associate Dean at C-MU, deserve our special thanks. So do Fiona Stevens and Brian Devine of Academic Press, whose support and counsel were invaluable.

PREFACE

The sciences that attempt to discover the causal structure behind statistical data include the study of the causes of disease and of the circumstances that spread particular diseases. They include aspects of genetics, archeology, chemistry, psychiatry, psychology, and all of the social sciences. There may not exist a scientific subject entirely free of statistical models and causal inferences drawn from them.

Nowhere is this influence more pronounced than in the social and behavioral sciences, from econometrics to sociology to political science to educational research, and nowhere are scientific conclusions of greater importance. A just society must make its collective decisions more or less rationally, and decisions cannot be rational unless they are informed. Few things are more evil or more dangerous than a society in the grip of fundamental and irrational dogmas, and few defenses seem more powerful than free and vital social sciences. We rely on social scientists to inform us about the effects of economic policies, about the effects of our educational system, about how we feel about ourselves and others, about the causes of behavior patterns we abhor, whether delinquency or racial discrimination, about the effects of our international policies, about the consequences of our attempts at social engineering, whether in birth control or in school integration. In a thousand policy decisions, large and small, made each day, the weight of social scientific opinion helps decide the balance.

Our courts, our military, our educators, our legislators may rely on the social sciences in one way or another, but reliance is not the same as trust, and the social sciences are not trusted. Unlike the natural sciences, the social sciences have little or nothing of a well-established foundation that can be applied and expanded to encompass more and more of social phenomena. There is no Isaac Newton in sociology, political science, or economics, nor should we expect one soon, or perhaps ever. The social sciences are therefore committed to the attempt to understand the causes of social phenomena without the aid of well established general theories, as if to do engineering without physics. It is no one's fault that we lack general theories that can reliably guide us to an understanding of social phenomena. It is not for want of trying that we lack them. In the absence of such theories, social scientists in many disciplines do work that has an air of arbitrariness; they produce statistical models and draw causal conclusions without much hint of serious argument for their assumptions. It is only natural to attempt to fill the vacuum of theory with the stuff of methodology, but the result is usually unsatisfactory. Social scientists are buffeted by methodological dogmas, many of them held dearly for reasons long forgotten

or never known. Much of the methodology is unhelpful; it gets in the way of scientific progress, and makes a poor substitute for genuine scientific sensibilities.

Wanting a powerful and accurate theoretical framework, and with difficult problems and debilitating methodological principles, mistakes are bound to be made, whether in advocating conclusions that are false or in missing conclusions that are true and important. It is no wonder that the social sciences are not trusted, and not surprising that critics within and without have proposed that the enterprise be abandoned.

This book is written with four convictions. The first is that next to the limitations on experimentation, the foremost problems of quantitative social science are computational. They are problems that result not from the failure of social scientists to adhere to every statistical nicety, but from the fundamental inability of humans—any humans—to carry out large calculations rapidly and accurately. The computations have to do with the systematic search for the best alternative explanations of statistical data. The human limitation is one of a sequence of logical dominoes: Unaided, we humans simply cannot consider and evaluate many alternative theories of non-experimental or quasi-experimental data. In the absence of well confirmed general theories, we have no other way of justifying our causal explanations than by making a case that there are no better explanations to be had, and we cannot make that case unless we have the power to search through a good sample of the possibilities. Unaided, no one has that power.

The second conviction we share is that the fundamental problem of computational limitations can be remedied. Unlike many critics, we do not believe that the practice of quantitative social science should be abandoned. We think it should be aided, and that it can be. The increasing availability of powerful, inexpensive digital computers puts enormous computing power within the grasp of nearly every researcher. The principal aim of this book is to describe a computer program designed to aid social and behavioral scientists, and those in other sciences as well, in searching through a vast array of possibilities to find the best explanations for their data. The program is not a toy, but it *is* only an example. The program is adapted to a particular kind of statistical model, and even for that kind of model the program could be, and will be, developed much further than it has been. We believe that the program is only a tiny fragment of what can be done to overcome human limitations in discovering theory. We are proud enough of it, and a little exhausted from its construction, but it is only a harbinger.

Our third conviction is that to gain the advantages of computational aids social scientists must throw away some ill-founded convictions about proper scientific method. Above all, they must dismiss the prejudice that any theory discovered by a machine that looks at the

data is *ipso facto* inferior to any theory discovered by a human who doesn't.

Our fourth conviction is that academic boundaries must be trespassed. The work described in this book is part applied statistics, part artificial intelligence, and part philosophy of science. These are distinctions in professions, not in subject matters, and we make no apology for mixing them so thoroughly. The mixture makes the reading a little tougher, the references much wider, and some parts of the argument a little more difficult for anyone enmeshed in a particular discipline, but the boundaries of the disciplines are not the boundaries of the issues. In intellectual life, good fences don't make good neighbors.

1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100

PART I

Artificial Intelligence and Non-Experimental Science

1. THE PROBLEMS OF SCIENCE WITHOUT EXPERIMENTS

1.1. THE LIMITS OF EXPERIMENTATION

Modern science and experimental methods grew up together, and we cannot very well imagine empirical science without experimentation. Even so, much of what we want to know about our world cannot be answered by experiments, and much of what we know (or think we know) is not directly based on experimental studies. The limitations of experimentation are both practical and ethical. For practical reasons we cannot do real experiments with the economies of nations or with the arrangements of galaxies. For ethical reasons we cannot do all sorts of experiments on the causes of disease (on smoking and cancer, for example), or on causes of social conditions; neither the world nor conscience will let experimenters randomly assign infants to experimental and control groups for the purpose of studying the effects of poverty on delinquency.

Faced with urgent needs for all sorts of knowledge, and with stringent limitations on the scope of experimentation, we resort to statistics. In the 19th century there was a rapid development in the study of statistical data to answer causal questions. Statistical comparisons were made in epidemiology, to determine the effects of blood letting or of new surgical procedures, for example, and to attempt to estimate the effects of occupation on mortality. Later, non-experimental statistical data were used to study heredity, the causes of differences in individual human abilities, and causes of economic change. By the early years of the 20th century non-experimental statistics had made inroads into medicine, biology, psychology and economics.

Today, statistical methods are applied to almost every public issue of any factual kind. The meetings of the American Statistical Association address almost every conceivable matter of public policy, from the safety of nuclear reactors to the reliability of the census. The attempt to extract causal information from statistical data with only indirect help from experimentation goes on in nearly every academic subject, and in many non-academic endeavors as well. Sociologists, political scientists, economists, demographers, educators, psychologists, biologists, market researchers, policy makers, occasionally even chemists and physicists, use such methods.

Non-experimental statistical studies are only one way to get information that helps to

decide questions of health or public policy. We could instead resort to fundamentalist revelation, or Marxist theory, or look entirely to political expediency. If we try to make science reach a little beyond our power to experiment, it is because we think such methods extend the scope of rationality, and because we wish to make social and individual decisions as rationally as we can. Even so, their reasonableness is not beyond dispute.

Controversy about the use of statistical methods to extract causal information from non-experimental data is as old as the methods themselves. The controversy continues today, and it has many sources. Part of it may be no more than the continued resistance to quantitative methods of many of the practitioners of the "softer" sciences. Part of it is baldly political. The same methods that are used to study the spread of disease or the causes of economic change have also been used to argue for some very unpopular conclusions. Part of the opposition to causal inference in non-experimental research is based on misconceptions about science. Critics unfamiliar with the history of science, or with practice in the natural sciences, sometimes make naive demands on applied statistics that are not met by even the most advanced of our sciences. But the most interesting and most important opposition derives from a justifiable sense that a good deal of applied statistics looks more like pseudo-science than like science. In many statistical studies in the social sciences, equations may be written down and references cited, but little or no justification is given for the assumptions that are made, and the hypotheses put forward are not tested, and no predictions of any consequence are derived. Indeed, as we will see in later chapters, some of the quantitative work that appears in even the most prestigious of social science journals is disappointing, and in some cases even appalling. We will discuss a major social experiment on criminology, supported by millions of dollars of tax funds, conducted without any recognition of the difference between testing a theory and estimating a parameter. Any critical reader of the quantitative social sciences can find an abundance of studies that are equally deficient. The important question is not whether much of this work is poorly done, but why it is, and whether it need be.

1.2. THE LIMITS OF HUMAN JUDGEMENT

It is important to recognize the difference between practice and principle. If much of the quantitative work that aims to extract causal conclusions from non-experimental data is not very good, then that might be either because of reasons of principle that cannot be overcome, or because of individual lapses that can be overcome. The news is not very encouraging: some of the obstacles to good statistical modeling may be very difficult to overcome, no matter how fine our intentions may be. There are two distinct sources of discouragement, one from psychology, the other from elementary considerations of

combinatorial mathematics.

1.2.1. The Psychological Problem

About thirty years ago, Paul Meehl [75] compared human performance obtained, on the one hand, by psychologists using clinical interviews and, on the other hand, by simple linear regression of the predicted attribute on any statistically relevant features of the population. He found that on the average the predictions obtained by linear regression were never worse, and were usually better, than those made by clinicians. Meehl's conclusion is one of the best replicated results in psychology. If the issue is selecting police officers who will perform their tasks satisfactorily, or graduate students who will do well in their studies, or felons who will not commit new crimes when paroled, a simple statistical algorithm will generally do better than a panel of Ph.Ds. These sorts of results are not confined to clinical psychologists. A growing literature on "behavioral decision theory" has found that people, even those trained in probability and statistics, perform well below the ideal in a variety of circumstances in which judgements of probability are called for [54]. One of the most resilient errors is the tendency to ignore base-rates, or prior probabilities for various alternatives, and instead to judge the probability of an alternative entirely by how likely that alternative renders the evidence.

The intellectual skills that are required to find, recognize and establish a good scientific theory have many of the features of problems for which it has been shown that human decision making is less than optimal. There are a few studies which bear a little more directly on the capacity of humans to interpret scientific data. Studies of physicians' diagnostic behavior, for example, have shown that few alternative diagnoses are entertained, and that evidence that is irrelevant to a preferred diagnosis is often erroneously taken to confirm it. Historical studies suggest much the same thing. Thus in recent years philosophers of science have been stung by the historical criticisms of writers such as Thomas Kuhn, who argue that the history of science does not fit very well with philosophical theories of scientific reasoning. The conflict is perfectly understandable if one realizes that philosophical methodologists have been concerned to characterize ideal, normative modes of reasoning, and if one supposes that the scientific community in any time and subject area typically falls well short of any such ideals.

One major use of applied statistics is in economic forecasting. Forecasts are made in many ways: by the judgement of a human expert, by a variety of statistical time-series methods, some of which require a good deal of judgement on the part of the user and some of which do not, and by econometric methods that attempt to consider not only time-series,

but also causal factors affecting the variable to be predicted. The literature on the comparison of methods is large and complex, and especially controversial since forecasting procedures may involve large financial stakes. It seems fair to say, however, that there is considerable evidence that simpler procedures do as well or better than more complex procedures, and that statistical procedures requiring considerable judgement on the part of the user do little if any better than more fully automated procedures. There is little evidence that explicit consideration of causal variables significantly improves forecasting (see Makridakis, et al. [71] and Kmenta and Ramsey, [57] for some recent comparisons and for references to the literature). What this suggests is that the causal theories produced for economic forecasting are not so good as we might wish.

Considerations such as these have led David Faust [24], a psychologist, to the conclusion that science itself is by and large too difficult for human cognitive capacities, and even Meehl takes this thought seriously. But there is a much more powerful combinatorial argument that leads to similar misgivings.

1.2.2. The Combinatorial Problem

The aim of science, whether physics or sociology, is to increase our understanding by providing *explanations* of the phenomena that concern us. The most common form of explanation in the sciences is to account for why things happen as they do by appealing to the causal relations among events, and by articulating generalizations about causal relationships. Causal claims alone are often insufficiently precise to give us the power to test them or to make important predictions from them. To gain that power, we usually embed causal claims, when we can, in a system of quantitative relationships. In considering non-experimental data, the quantitative relationships are often in the form of a *statistical model* of some kind.

When we consider statistical models¹ in the social and behavioral sciences, we find everything from very small models in educational research with as few as six or seven measured variables, to econometric models containing several hundred variables. We can think of the causal part of any such theory as given by a **directed graph**, with the vertices of the graph representing the variables, and each directed edge in the graph representing a causal influence of one variable upon another. Such graphs look like arrow diagrams. For example:

¹The practice throughout statistics and much of the sciences is to term any specific theory a "model" and reserve the term "theory" for only the most general and sweeping of hypotheses. We will use "theory" and "model" interchangeably.

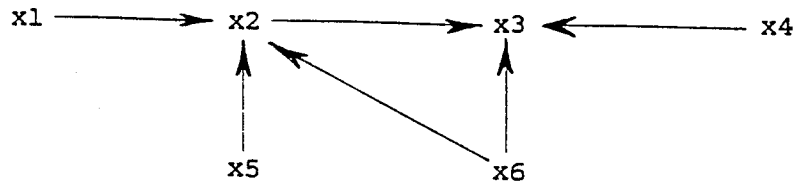


Figure 1-1:

Each distinct way of arranging a set of variables into a directed graph represents a distinct model of the causal relationships among the variables, and thus an alternative causal theory of the data. The point is that even with a small collection of variables, the number of distinct possible causal arrangements, and thus the number of distinct, alternative causal models of the data, is astronomical. If we consider all of the *a priori* possibilities, then for each pair of variables A and B, there are four possible kinds of connection: A can have an effect on B but not B on A, or B can have an effect on A but not A on B, or A and B can each have an effect on the other, or, finally, A and B can each have no direct effect upon the other. The number of distinct possible causal arrangements of n variables is therefore 4 raised to the power of the number of pairs of variables. Thus with only six variables, there are 4^{15} different possible directed graphs or causal models. When we have it in our power to experiment, we are able to arrange circumstances so that we know most of these possibilities are excluded, and we can focus on whether a particular causal dependency does or does not occur. But without experimental control, the problem of determining the correct causal structure is far more difficult.

A social scientist or psychologist or epidemiologist or biologist attempting to develop a good statistical theory has a lot of difficult tasks. He or she must choose what to measure and how to measure it, and worry about sampling technique and sample size. The researcher must consider whether the variables are multinormally distributed, or have some other distribution. He or she must worry about whether measures of a variable in one individual or place or time are correlated with measures of that same variable in another individual or place or time. The researcher must consider whether the relations among the variables are linear, or non-linear, or even discontinuous. These are demanding tasks, but there are a variety of data analytic techniques to help one along and to test hypotheses concerning these questions. Suppose the investigator has passed these hurdles and has arrived, however tentatively, at the usual statistical modeling assumptions: the relationships are linear, or close enough to linear, the distribution is multinormal, or nearly so, there is no autocorrelation, or at least not much. Assume that the investigator has covariance data

for six variables. The troubles that now must be faced make the data analysis problems seem almost to pale in comparison. There are 4^{15} alternative possible theories of the causal dependencies among the six variables, and only one of those theories can be true. How is the researcher to find the one needle in the enormous haystack of possibilities? Perhaps the goal need not be to find the one unique correct theory, but only to reduce the possibilities to a handful, or to find a close approximation to the correct model. How is even that to be done? The usual thing to say is that the investigator must apply his or her "substantive knowledge" of the domain, but this is mostly whistling in the dark. To get the number of possibilities down to manageable size will take a lot more substantive knowledge than we usually have about social or behavioral phenomena. Suppose, for example, that one is quite certain that the causal relations do not form cycles. That is, there is no sequence of directed edges in the true graph of the causal relations that leads from one variable back to that same variable. Then the number of alternative causal models consistent with that restriction is still more than three million. Suppose, what is in fact rather unusual in many social science studies, that the variables are ordered by time of occurrence, and the researcher therefore knows that variables occurring later cannot be causes of variables occurring earlier. There are still $5!$ or 120 alternative models.

Now repeat the same sequence of calculations when there are twelve variables in the data set. With no restrictions imposed, there are 4^{66} alternative causal models, only one of which can be true. If the researcher knows that there are no cyclic paths in the true model, the number of alternatives is still beyond the astronomical: 521,939,651,343,829,405,020,504,063 (see Harary, [41]). If the researcher is lucky enough to be able to totally order the variables and knows that later variables cannot cause earlier variables, the number of alternatives is reduced to $11!$ or a mere 39,916,800.

These counts are conservative. They do not include the possibility, which every researcher considers, that some part of the correlations is due to unmeasured variables that have effects on two or more measured variables. Including such possibilities enormously increases the numbers. What can the investigator possibly do in the face of such an enormous number of possible causal models, no more than one of which can be correct? In practice, even the best researchers usually take a wild guess, or indulge their prejudices rather than their knowledge. One, or at most two or three, causal models is suggested, appropriate equations are written down, parameters are estimated and the model is subjected to some statistical test. If the test is passed, the researcher is happy and journal editors are happy and the thing is published. No one likes to mention the millions, billions or zillions of alternatives that have not been considered, or the dark and troubling fact that among those alternatives there may be many which would pass the same statistical test quite as well or even better, and which would also square with well-established

substantive knowledge about the domain.

The practice of model building in the social and behavioral sciences looks very much like irrational judgement under uncertainty. The very phenomena that Kahneman and Tversky and many others describe in experimental subjects seem to be exemplified in the practice of social science. According to Bayes' theorem, the posterior probability of any particular model should take into account both the likelihood of the evidence on the model, and the prior probability of the model. Much of current practice is to hit upon a model, somehow, and to advocate the model if it passes statistical tests, even when the tests are not powerful against all alternatives. In effect, those who follow this practice are ignoring the prior probabilities, and judging a model to be good enough to accept if it yields a sufficiently high likelihood. Other workers may do much better than this, and attempt to carry out a serious search for alternative explanations of the data. But in view of the vast numbers of possible models, we should wonder how anyone *could* carry out an adequate search.

1.2.3. Some Examples

These arguments are not just idle combinatorics. The issue affects almost every piece of applied statistics addressing non-experimental data. Consider just a few recent examples:

1. *The Census*

The national census is certainly not free of error. Some people are sure to be missed, and those likeliest to be missed are often the least advantaged in our society: those without fixed abodes, those who live in isolated areas, those who live where census workers would as soon not go. Since census statistics are used to apportion all sorts of things, from votes to benefits, who is missed where is of considerable practical and political importance. The statistical task is to estimate the actual undercount in various places, given the actual count, and given estimates of undercounts based on matching studies.

For the 1980 census, the Census Bureau produced 29 matching studies (Post-Enumeration Program, or PEP studies), based on a variety of alternative assumptions. Using one of these PEP studies, Ericksen and Kadane, [23], investigated a number of alternative linear models in which the actual undercount depends on geographic, social, procedural or demographic variables, and the PEP estimates are a function of the true undercount. The distributions are assumed to be normal. They estimate the undercount by regression, and argue that the

undercount is robust over the alternative models they consider.

Ericksen and Kadane's work was criticized by Freedman and Navidi, [27], on several grounds. One of the criticisms lodged is that Ericksen and Kadane do not justify their assumptions. The point of the criticism is that allegedly, for all anyone knows, alternative PEP series have the correct assumptions, and alternative models are correct and would lead to different estimates of the undercount, or to no estimates at all (if the undercount is represented by an unidentifiable parameter). Kadane [53] replies that (among other things) there is a justification for the PEP series selected, that the conclusions he and Ericksen obtain are reasonably robust, and that the undercount estimate on any reasonable set of assumptions would be better than the Census Bureau's *de facto* estimate, which is that the undercount is zero. It is clear that a major issue in this dispute concerns whether an adequate search can be made for plausible alternative linear models, and whether estimates of the undercount are reasonably robust over that collection of alternatives.

2. Criminology

McManus, [74], (reported in Leamer, [59]) investigated the deterrent effect of capital punishment by regression on several subsets of variables. He obtained importantly different results according to the independent variables selected. It is, however, unclear why the sensible thing to do is not obvious: include all of the variables. Even so, many of the independent variables considered are undoubtedly correlated, and alternative assumptions about their dependencies can be expected to again give different estimates of the deterrent effect.

In one of the largest social experiments ever carried out, described by Rossi, et al., [82], newly released felons in Georgia and Texas received unemployment payments for a period of six months after their release. Legal requirements prevented those receiving payments from working. The rearrest rate was the same for the Texas group and for controls who received no payments, and nearly the same for the Georgia group and controls. The experimenters concluded that payments did decrease recidivism substantially, but that unemployment increased recidivism, and the two effects exactly, or almost exactly, canceled one another. They elaborated their conclusions in a path model. In effect, the experimental design failed to control for a relevant variable, unemployment, and thus left open many alternative explanations of the results. Zeisel [110], who served on an advisory committee for the experiment, vehemently objected to the conclusions drawn by Rossi, et al. In his opinion the experiment established the obvious: payments have no significant effect on recidivism. Zeisel proposed a simple alternative path model to account for the data.

3. Head Start

Data on socio-economic status variables and achievement and ability scores for Head Start participants and comparable non-participants has been much studied and discussed, and has had an effect on government policy. A principal difficulty with many analyses in the literature, beginning with the Westinghouse evaluation of the summer Head Start program, is that researchers have failed to carry out any systematic search for alternative models. Jay Magidson, [69], in a study of the summer Head Start data, put the issue this way:

The problem we face is that there is an infinite number of ways to formulate a causal model, and it is not a straightforward matter to determine how to go about doing it, particularly when the causes are unknown and/or unobserved. It is important for researchers to formulate not one but many models so that they can determine whether their conclusions may differ if they accept a different set of assumptions. It is also important to follow some general guidelines in building (formulating) models when the researcher has limited information about the causal process...

1.3. THE ARTIFICIAL INTELLIGENCE SOLUTION

The problem we have described seems to us both urgent and difficult. In its way, the problem is far more difficult than that of getting people to do proper data analysis, to perform tests for linearity, for example, or for autocorrelation, or to select the most powerful test statistic. There are procedures for diagnosing erroneous statistical assumptions, and work has been done towards automating such procedures, especially in the context of regression analysis. Even so, the chief difficulty will usually remain: the number of alternative theories is astronomical, and it is beyond anyone's capacity to analyze and to test any more than an insignificant fraction of the possibilities.

An ideal Bayesian solution would be to impose a prior probability distribution on all of the alternative statistical models that are consistent with prior knowledge (or with prior knowledge and some simplifying assumptions), compute the likelihood of values of an appropriate statistic as functions of the model parameters, obtain the evidence, and with the likelihood and the prior probability distribution, compute the posterior distribution.² One trouble with this solution is that the sort of evidence we have may be insufficient to bring about reconciliation of agents with different priors. Leamer, [59], finds, for example, that representing the views of Keynesians and monetarists by different prior probability distributions, and conditioning on time series evidence doesn't do much to change initial

²This strategy was suggested to us by our colleague, Jay Kadane.

prejudices. Another trouble is computational. We don't possess a general procedure for quickly computing posterior probability distributions from arbitrary prior distributions and arbitrary evidence. Perhaps some day such a procedure will be feasible, but what do we do in the meanwhile?

Two procedures are in common use to help with this problem, and while they have their virtues, neither is adequate. One procedure is exploratory factor analysis, which is sometimes very useful in suggesting that groups of measured variables cluster together and may therefore have a common cause. But exploratory factor analysis does not consider the variety of possible causal relationships. As Blalock [8] points out, factor analytic procedures do not consider the possibility that measured variables may have direct causal effects on one another, nor do they allow that measured variables may have direct effects on unmeasured variables, nor do they consider the different possible causal relationships among the unmeasured variables. Factor analysis procedures rule out a variety of realistic possibilities *a priori*.

Another procedure commonly used to help search for alternative models is the formation of *nested sequences* of models. The idea is really quite old, and goes back at least to Harold Jeffreys' *Scientific Inference*, [49]. There are many different technical forms of the idea, but the basic notion is that one starts with a simple theory having only a few free parameters, or, equivalently, with a theory that postulates only a few causal relationships among the variables it considers. One then introduces a new free parameter into the model, or new causal connection among the variables. Sometimes the parameter freed is chosen by a formal procedure. Some statistical test is used to compare the new model with the model that precedes it, and the process is continued until eventually the statistical test is failed. A procedure like this is carried out automatically by the LISREL program, which is widely used for certain kinds of statistical modelling. While the general strategy is admirable, and the TETRAD program described in this book has an analogous architecture, in practice there are severe difficulties. The trouble with the procedure is that there are generally far too many possible nested sequences to be explored in this way. The results obtained depend on the order in which the parameters are freed, and the formal procedures for choosing which parameters to free are too fragile, and tend to overlook the best options. Some of these difficulties were noted by those who introduced some of the technical procedures (see Sorbom, [97] and Byron, [13]).

This book proposes a new solution, or more accurately a partial solution to the problem of searching for alternative models. That solution has two parts. One part is to consider formal aspects of *scientific explanation* in comparing alternative models. The idea is simply that we should prefer those models that offer the best explanation of our data, and that

important aspects of explanation can be represented by formal, mathematical relationships between a model and the data it is to account for. "Explanation" is not a methodological notion that plays a major role in statistics, and the theory of scientific explanation upon which we rely is not contained within any of the familiar schools of statistics. In thus broadening the methodological viewpoint we do not intend to neglect statistical inference where it is applicable and useful, but we do intend to make use of an understanding of scientific explanation that is familiar to philosophers and historians of science, and that is common in the natural sciences.

The second part of our partial solution is to use artificial intelligence techniques to help search for models that will provide the best explanation of our data. The very idea owes a great deal to Herbert Simon, [92], who also contributed to the understanding of connections between causal relationships and multivariate analysis. Simon takes one of the hallmarks of artificial intelligence to be *heuristic search*. A heuristic search is a computer procedure that applies plausible steps to hunt through an enormous space of alternatives to locate the best (or a collection of the best) alternatives for some purpose. What makes the search *heuristic* is that the procedure does not *guarantee* that the outcome will be the best alternative or will be a collection that includes the best alternative, but the procedure will typically do so, or come rather close. The search procedures may in various respects be less than rigorous, and may not always deliver the optimal solution. But they will typically yield outcomes that are good enough, and that are better than can be obtained without heuristic search. Simon calls the strategy "satisficing." It amounts to settling for what is feasible and good enough, rather than insisting on what is optimal but infeasible. A philosopher, Wilfrid Sellars, puts the same point in a simple imperative, upon which we will have several occasions to rely: do not let the best be the enemy of the good.

Simon's idea is that scientific discovery is a kind of heuristic search through alternative hypotheses or theories. If that is so, then there is a formal structure to scientific discovery, and if we can get a grip on that computational structure, then computer programs can make scientific discoveries. And indeed they already have. In chemistry, in logic and in geometry, artificial intelligence programs have made discoveries of various kinds, and Simon [93] and his associates have written a series of programs that simulate many of the discoveries in the history of science.

The aim of artificial intelligence programs need not be to do things in quite the way that humans do them, especially if humans do the thing rather poorly, and the computer can be made to do it better. Humans do not do large sums very accurately; we do not want our pocket calculators to simulate our own inadequacies. Humans also seem not to be able to search through the space of alternative statistical models very adequately. Our claim is that

a computer program, using heuristic search techniques, can help do it better. This book is a defence of that claim. Our argument is based on an actual program, TETRAD, a copy of which is included with this book. The program is intended to assist in the search for adequate statistical models of a very general kind, including factor analytic models, path analytic models, structural equation models, and many other varieties of linear statistical theory. They do not include statistical theories that cannot be put in additive form. With all modesty, we believe that a TETRAD user can learn more in an afternoon about the existence and properties of alternative linear models for covariance data than could an unaided researcher in months or even years of reflection and calculation. The TETRAD program is offered as a useful working tool, but it is also offered as an example. We think the same very general ideas about heuristic search can be applied to develop artificial intelligence aids for other kinds of statistical models, and they ought to be.

The notion of applying artificial intelligence techniques to problems of statistical model specification is not really novel. Several programs have been developed or proposed that apply artificial intelligence techniques to statistical modelling. Systems have been developed to test automatically the adequacy of common modeling assumptions, such as linearity and normality, and to suggest transformations of variables or other changes where these assumptions fail. (See Gale and Pregibon [30] for a review). A great deal is known about statistical diagnostics, and the aim of these programs is to make that information, or appropriate conclusions, automatically available to the inexpert. One example is the REX program developed at Bell Labs by Gale and Pregibon. The program is designed to assist in regression modelling, and will actively transform variables (to their logarithms for example) to satisfy regression assumptions.

In principle, computer aided search for causal relations has important applications beyond aiding researchers in finding better explanations. Fully automatic programs of this kind can function as inference modules in robotic systems. More immediately, fully automatic search for causal relations holds the promise of unlocking large data bases. In medicine, social science, astrophysics, in fact almost every domain, we have managed to collect more data than we have time and power to analyze. Potentially valuable causal knowledge is effectively locked up by the sizes of the data bases. Computer discovery procedures that are fully automated and reasonably reliable seem the only hope for making real use of the information we continue to acquire and store.

The problem of automatic computer search for causal relations has barely been scratched. Perhaps the best known piece of work in this line is Blum's [11] RX program. The RX program uses prior knowledge about a particular domain, rheumatoid diseases, together with a statistical package and a large data base. The variables in the data base are indexed

by time, so that a variable with a later index cannot be a cause of a correlated variable with an earlier time index. The program looks for correlations that may be due to causal relations. When a correlation is found, the relationship is tested and compared with the program's current knowledge to determine whether the correlation may be due to common causes acting on the variables. If not, a causal conclusion is added to the knowledge base and the program continues. The statistical analyses are restricted to multiple regression models, and therefore a great many alternative causal explanations of correlations are never considered.

Unlike Blum's program, the program we describe in this book, TETRAD, is not fully automatic and does not proceed without the active engagement of the user's judgement.³ But neither is the TETRAD program confined to searching for multiple regression models. Instead, the program is intended to help the user search through the vast space of alternative linear causal models that might explain a body of covariance data. But the very idea of the TETRAD program is really nothing more than a continuation of the tradition of work on automated discovery carried out by Simon and Buchanan, [66], Blum and others.

Conventional statistics constantly reminds us that there are two sorts of errors a theorist can make: he or she can reject the correct theory, or fail to reject a false theory. Perhaps the likeliest reason for advocating an incorrect theory is simply that the correct model has not been thought of or investigated. Both sorts of errors might be reduced by helping investigators to search more adequately for promising explanations of their data. The TETRAD program is designed to do exactly that.

Most of the ideas in the TETRAD program have a long history in many different disciplines: philosophy of science, psychometrics, statistics, and elsewhere. Some of that history is described in a later chapter. Other chapters of this book describe some of the philosophical issues that surround artificial intelligence, causal modeling, and the explanatory strategy of the TETRAD program. They also describe how TETRAD works, and illustrate the use of the program on a variety of real and simulated cases.

³We think the ability to infer causal relations without user interaction is a considerable advantage of Blum's program, particularly in searching large data bases, whether in medicine or in social science. We have developed pilot programs, which we will not describe in this book, that more fully automate the TETRAD procedures.

2. THE CASE AGAINST CAUSAL MODELING

2.1. THE CRITICAL REACTION

Causal modeling of non-experimental data has been controversial since its beginnings, and it is no less controversial today. The controversies are very philosophical, and they involve fundamental differences about what makes for science. The social and behavioral sciences are inevitably compared with the natural sciences, and disputes about methodology in social and behavioral science are often essentially disputes about what it is that has made the natural sciences so successful in understanding, predicting, and controlling the physical world.

Paul Meehl, [76], is fond of pointing out the discrepancies between the methodological procedures of chemists and physicists and biologists, on the one hand, and the methodological practices of the "soft sciences" on the other hand. Most natural scientists would be altogether puzzled by what goes on in quantitative sociology, or social psychology, or econometrics, and they would be unlikely to recognize some of it as science. Some of it isn't science save in name, however much it may try to be.

The essential question we shall pursue is whether, in the many criticisms of causal modelling *practices*, there are good reasons to think that the entire endeavor is suspect, or whether the valid criticisms are only evidence that particular social and behavioral scientists don't adequately understand what science requires. That the enterprise is often badly conducted is no sufficient argument that it is inherently unscientific, or that it could not in principle be done well.

Our discussion will also consider just *which* of the criticisms of causal modelling are well taken, and which derive from misperceptions and misunderstandings. Our view is that there are many well founded criticisms of particular studies, and of various common practices in causal modelling, but none of the criticisms gives reason to think that the very idea is mistaken. We will also argue that the critics are often nearly as confused as those they criticize, and that many of the principled criticisms of causal modelling derive from mythology about the natural sciences. Our view is that there is nothing philosophically or methodologically wrong with applying statistical models to non-experimental data, or with the attempt to uncover causal structure. But doing it well, so that the result contributes to the progress of knowledge, is very hard, and there are many ways to go astray.

We will consider in turn the following criticisms of causal modelling:

- Causal modelling involves a mistaken, or incoherent, conception of causal relations.
- Theories with latent variables should be rejected on methodological or semantic grounds.
- Only experimental data can contribute to our knowledge of causal relationships.
- Those who advocate causal models do not, and presumably cannot, make a case for the assumptions of their models.
- Linear causal models should be rejected because they are always literally false.

2.2. MAKING SENSE OF CAUSALITY

Some social scientists, prefer to avoid causal language. They assume it is enough to give a set of structural equations, estimate the coefficients of the model, and perhaps subject the system of equations to a statistical test or tests. Some statisticians regard causal inferences drawn from non-experimental data with the assistance of statistical models as an abuse of their subject.⁴ These harsh judgements seem to spring from a vague sense that causal talk is unscientific or "metaphysical", or at best a gratuitous and unclear addition to the much clearer system of equations of a structural model.

In fact, the very logical structure of most social science models requires mathematical structure beyond that of a system of equations. Econometricians distinguish exogenous from endogenous variables, others distinguish independent from dependent variables. These distinctions are not given by the equations themselves. It makes no difference to the algebra which variables are written on the left hand side and which on the right hand side of the equality sign. When social scientists introduce such distinctions they are providing, in addition to the equations, a partial ordering of the variables. That partial ordering has a natural causal interpretation.

The application of most social science models depends on drawing causal conclusions from them. That is because we are usually interested in applying such models in situations where we have the possibility of changing some features of the social or economic system, and our concern is to know what other changes will result if we do so. This sort of

⁴For some more optimistic recent discussions of statistics and causal inference, see Holland [45], and Glymour [36].

application is routine with econometric models of national economies. Parameters representing policy variables are run with hypothetical rather than actual values in an attempt to predict the effects of changes in policy. Again, with models of the Head Start data, we are concerned with whether or not Head Start participation *causes* improvement in school and test performance. In studies of the American occupational structure, we may be interested in whether the educational system *causes* social mobility. In criminological studies we are often concerned with whether specific penalties (e.g., the death penalty) *deter* certain crimes. To deter is to cause not to happen. Correlations and regression coefficients don't give us that information. Only causal conclusions will do so.

There are some contexts in which social science models are of practical importance even without causal conclusions. In estimating the undercount in the Census, for example, we are not principally interested in drawing a causal conclusion. But more often, we want causal knowledge, however much our language may disguise that fact. Causal relations are either asserted or presupposed in almost every circumstance in which the application of social theory leads to a *counterfactual* assertion, that is to an assertion roughly of the form "If A had happened, B would have happened". Causal conclusions are usually implicit when *future conditionals* are asserted, for example in sentences such as "If the death penalty is enforced for murder, fewer homicides will occur." If social science could never lead to these sorts of conclusions (or their denials) we would not have much practical use or hope for it⁵

The idea that causal discourse is somehow unscientific is rather wild. The determination of causal relations and causal structure remains a principal part of the enterprises of physics, chemistry, biology, engineering, and medicine. Causal talk is used without a qualm in many scientific papers in these subjects, and in every laboratory. Why is it improper then in social science? It is sometimes claimed that causal talk is "meaningless," but that is itself sloppy expression that cannot be taken seriously. We use causal talk all the time in everyday life, with pretty fair mutual understanding. We have excellent formal semantical theories of causal discourse (see Lewis, [62] and Cross, [20]). In innumerable cases we know how to determine causal relationships. What more could be required for intelligibility?

⁵For a careful discussion of the confusions about causality in the recent methodological literature in economics, see D. Hausman [42]

2.2.1. Interpreting Causal Claims

Many philosophers have attempted to explicate the notion of causality in terms of the notion of probability (see Suppes, Skyrms and Good, [103, 96, 37]). A probabilistic characterization of causality has been introduced into econometrics by Clive Granger, [39], and developed by others (see Sims [94]). We do not assume that causal relations can be reduced to probabilistic relations of any kind, but neither do we contradict such assumptions. Instead, we assume that the theorist imposes certain connections between causality and probability. The most important of these connections is that, *ceteris paribus*, correlations are to be explained by causal relationships of one or another kind.

We assume that a causal claim implies a functional dependence, although not every functional dependence is a causal dependence. If A is a cause of B, then B is a function of A and (quite possibly) other variables. If B changes, then A must have changed if the other variables of which B is a function did not change. So we say that if A is a cause of B, then *ceteris paribus*, a change in B must be accompanied by a change in A. Or, put another way, if A causes B, then *ceteris paribus* if a change in A had not occurred, a change in B would not have occurred.⁶

Mathematically, the representation of causal relationships in combination with linear structural equations systems is very simple. The causal relations are represented by a labeled directed graph, and the directed graph uniquely determines the structural equations of the statistical model. But mathematical representation is the easiest part; the hard part concerns what inferences are to be drawn from causal claims. One of the inferences to be drawn from the claim that A causes B is that, *ceteris paribus* if B has changed then A must have changed. In the context of causal modeling the claim is ambiguous. There is a strong version, namely:

For any member of the population, *ceteris paribus* if the value of B for that individual has changed then the value of A for that same individual has changed.

There is also a weaker interpretation of the claim that variable A has an effect on variable B:

For any two members of the population, *ceteris paribus* if they differ in their

⁶ It might be claimed that we have it backwards, and that A causes B means that other things equal a change in A is accompanied by a change in B. But if A causes B does imply that B is a function of A and other variables, then (unless the function is one-one) a change in the value of A can occur without any change in the value of B, as in $B = A^2$. If the relations are linear then the relation is of course symmetrical, and if A causes B then, *ceteris paribus*, if A changes B will change.

values of B then they differ in their values of A.

The two versions are not the same, and one can be true when the other is false. The first sense presupposes that if feature A is a cause, then it makes sense to talk of changing A for particular individuals. Sometimes that doesn't make sense: for example, it doesn't seem to make sense to talk of substantially changing the genetic structure of one and the same individual. Some writers, such as Holland [45], contend that only when one is prepared to assert the strong claim is one really talking of causality. Terminology aside, it is important to recognize that the weaker claim might be correct even though the stronger one fails.

A related consideration is that causal claims may be true only historically. Lieberson [64] points out, correctly, that many causal processes are not reversible. So if values of B came to obtain historically because of what appears to be a linear dependence of B on A and changes in the value of A, a reversal in the value of A might not bring about a concomitant reversal in the values of B. When that occurs the dependence of A on B is not in fact linear.

These points are really cautions rather than objections to the notion of causality or to the search for causal explanations of non-experimental data. They remind us that in understanding causal claims we must pay attention to what is being talked about. Lieberson enters another caution of a similar kind: If one knows that A causes B, say, one cannot conclude and should not in general conclude that a change in a third variable, say C, that has not hitherto varied in the population or sample, will not cause a change in A or in B or in both.

All of these cautions are valuable, but there is no reason to suppose that they must necessarily bring trouble for causal modelling. Anyone considering genetic causes of phenotypic traits, for example, will be unlikely to confuse the weak and strong *ceteris paribus* conditionals. We can consider what differences in genotype between individuals may produce differences in phenotypic traits, but we are not likely to confuse ourselves into thinking we can change the genotype of a particular individual. No one is likely to think that the claim that Head Start participation improves school performance implies that a sixteen year old who did not participate in Head Start will have his school performance improved by reversing that condition and entering Head Start at age sixteen. Few of us are likely to think that because nutritional levels did not vary in a sample of students obtained in a study of school performance, that major changes in nutrition would have no effect on school performance. The cautions are sound, and important, but there is nothing in them to make us abandon the enterprise of causal modelling.

2.2.2. Baumrind and the Concept of Causality

In the course of a criticism of a paper by O'Donnell and Clayton [79], arguing that marijuana use is a stepping stone to heroin use, Diana Baumrind mounts a vigorous attack on causal modelling. She has interesting substantive objections to their argument, and she proposes an alternative causal model of the data, which we will not consider here. The chief burden of her essay, however, is that causal modelling of non-experimental data can never justify drawing causal conclusions. It is hard to find in her essay a coherent argument for that thesis, but she does raise some interesting issues about the concept of causality.

Baumrind criticizes in particular the idea that causal relations consist in certain sorts of probabilistic relations, an idea which she (mistakenly) thinks is essential to causal modeling. The proposal, roughly, is that As cause Bs, where A and B are kinds of events, if the occurrence of an A event increases the probability of a subsequent B event, and there are no kinds of events C1,...,Cn, such that, conditional on prior events of kinds C1,...,Cn, events of kinds A and B are statistically independent. Her criticisms of probabilistic accounts of causal relations are these:

1. It is a "parochial model of causality shared by neither laypersons nor philosophers of science." (p. 1289)
2. Shultz [88] has performed experiments that appear to show that people of all ages prefer a "generative" model of causal relations to covariation.
3. The correct account of causation is generative: "The generative approach to causation refers to the notion that the cause *produces* the effect...Within the generative model the event called the *cause* acts to change or to produce the event called the *effect*" (p. 1291).
4. A parable: "The number of never-married persons in certain British villages is highly inversely correlated with the number of field mice in the surrounding meadows. Marital status of humans was considered an established cause of field mice by the village elders until the mechanisms of transmission were finally surmised: Never-married persons bring with them a disproportionate number of cats relative to the rest of the village populace *and* cats consume field mice. With the generative mechanisms understood, village elders could concentrate their attention on increasing the population of cats rather than the proportion of never-married persons. Note that although the correlation between marital status and incidence of field mice is *not* a joint effect caused by incidence of cats and is therefore a true association...the explanation that marital status is a cause of incidence of field mice is at best trivial, whereas the generative explanation that cats consume mice is valuable." (p.1297)

We cannot speak for lay persons, but we can speak for philosophers of science. Baumrind's claim is false. Several (rather more precise) versions of the probabilistic account of causality sketched above are advocated by some of the most prominent contemporary philosophers of science (For example Suppes [103], Salmon [84] and Skyrms [96]). We can speak *about* lay persons. What Schultz's experiments show is hard to say. No one should be surprised if in some circumstances lay people systematically violate probabilistic criteria for causal attributions. There is considerable evidence that lay, and even expert, judgements about probability and causality violate most every normative standard (see Kahneman, [54], and Faust, [24]).

Baumrind's characterization of the "generative" model of causality is not very helpful because it is not very clear. There are a few contemporary philosophers who analyze causal relations in terms of an unanalyzed oomph, but the closest thing to a clear and influential philosophical account that bears any connection to Baumrind's is the counterfactual analysis of causation developed by Mackie, [68] and by Lewis, [63]. On these accounts, A causes B if and only if A occurs and B occurs, and if A had not occurred then B would not have occurred. (In Lewis' theory a model theoretic semantics is developed for the counterfactuals that arise in this analysis of causation.) These accounts perfectly well allow increases in never-married persons to cause declines in the population of field mice.

What Baumrind really seems to mean by "generative" causation is that causal attributions should come with a specified *mechanism* by which the causal relation occurs. The intervening steps should all be spelled out, and verified if possible. That is a perfectly sensible request, but it is part of *methodology*, not part of the meaning of "cause." It is a request that can only be satisfied, in any subject, by the examination of correlations, experimental or otherwise, and the application of prior knowledge. Baumrind's parable is bootless. In the story, variations in never-married persons *do* cause variations in field mice, even if the causation is indirect, and nothing in the story prevents the use of covariance analysis on uncontrolled samples to discover that the intervening variable is the density of cats.

Lots of models derived from non-experimental data come with plausible speculations about the process by which the causal relation comes about. Thus the apparently astonishing conclusion (see Crain, [19]) that annual automobile safety inspections actually tend slightly to increase highway accidents rather than to decrease them is elaborated with possible mechanisms. For example, inspections may tend to give automobile drivers an erroneous view of the safety and reliability of their machines. There are obvious psychological mechanisms for the apparent effects of responses to some questionnaire items on

responses to other items. Blau and Duncan occasionally suggest mechanisms for the causal relations they argue for in *The American Occupational Structure*.

The importance of establishing the details of a causal process depends on the context. In many cases we are perfectly happy to know that plausible mechanisms can be conceived, and don't much care what the actual mechanism may be. In other cases, we may doubt a causal claim exactly because no plausible mechanism can be thought of. In still other cases, we may not have much doubt about the genuineness of the causal connection, but may be especially concerned to discover its mechanism of action. In these respects, sociological and psychological and epidemiological mechanisms are no different from chemical mechanisms.

2.3. CAUSES, INDICATORS AND THE INTERPRETATION OF LATENT VARIABLES

2.3.1. General Objections

Many linear models, including those for which TETRAD analyses are most useful, contain variables that have not been measured. Mathematically, models with such latent variables often specify that some of the measured variables are linear functions of one or more of the latent variables. Further, the latent variables are often given a causal interpretation--their variation causes variation in the measured variables--and this causal interpretation is represented in the directed graph of the model. Finally, latent variables are routinely given a title that carries some meaning with it: "socio-economic status", "industrial development", "authoritarian-conservative personality trait."

Realism and Anti-Realism

A long tradition in philosophy of science holds that what cannot be measured or seen directly is not worthy of *belief*. Philosophers (and they have included many distinguished scientists, such as Pierre Duhem) of this persuasion usually hold that theories that postulate unobserved objects and properties and relationships are useful and valuable *instruments* for prediction and control, for generating new experiments and for guiding the development of still further theories, but they deny that these virtues give grounds for belief. They distinguish belief from *acceptance*. We may have good practical reasons for accepting theories as good empirical predictors and as useful tools, they hold, but no reason to go further and actually believe our theories. The best contemporary defence of this anti-

realist conception is in Bas Van Fraassen's, *The Scientific Image*, [106].

Realist philosophers of science think otherwise. They think we can have as much, perhaps more, reason to believe in our theories as we do to believe in everyday claims made outside of science. They distinguish *credence* from *certainty*. Belief has to do with credence, and we can allot more or less of it to a proposition without taking that proposition to be *certain*.

Realists and anti-realists in philosophy of science agree that theories that go beyond observation are useful and inevitable, and they largely agree as well on what virtues make one theory preferable to another. What they disagree about, for the most part, is whether those virtues give us reason to believe our best theories.

In the social and behavioral sciences, and in applied statistics generally, there is a much more radical train of thought, which holds that in constructing theories *we should not talk about features that postulate unmeasured features or properties or entities*. This view is very different from the philosophical anti-realism we have just discussed, for it amounts to claiming that not introducing unobserved features is an overriding virtue of any theory; any theory that introduces unobserved features is inferior to any theory that does not.

These critics think that it is somehow unscientific to introduce theoretical causes that are not directly measured. For example Holland, [45] rejects the introduction of any causal factors that cannot be manipulated; and many econometricians and statisticians object to latent variables on the grounds that they do not give any novel predictions. The most famous modern rejection of latent variables is B.F. Skinner's [95], on the grounds that anything that can be predicted with latent variables can be predicted without them.

These critics have it exactly backwards. *The natural sciences are successful exactly because of their search for latent factors affecting the phenomena to be explained or predicted*. Newtonian dynamics and celestial mechanics, the theory of electricity and magnetism, optics, chemistry, genetics and the whole of modern physics would not have come to pass if natural scientists behaved as the critics of latent variables prefer. Gravitational force, electrical fluids and particles, electromagnetic fields, atoms and molecules, genes and gravitational fields, none of them could be directly measured or manipulated when they initially became part of modern science. Few of these notions were introduced initially because they predicted anything; they were introduced because they *explained things*. Newtonian theory explained Kepler's laws, and did so beautifully; Daltonian atomism explained (then controversial) regularities of chemical substances, such as the law of definite proportions; general relativity explained the anomalous advance of the perihelion of Mercury; Bohr's quantum theory explained spectral series of hydrogen. Of course these

theories did eventually lead to important predictions, but they did so exactly because people investigated the latent factors postulated by the theories, and sought laws about them and their properties, and sought to test those generalizations.

Now critics may grant that the introduction of latent variables has been essential in the natural sciences, but maintain that they are inappropriate in the social and behavioral sciences. It is hard to think of any convincing reason for this view, however, and we should demand *very* convincing reasons for any methodological strictures that contradict the methodology of our most successful sciences. Critics may doubt that statistical theories of nonexperimental data can *explain* in the same way that theories in the natural sciences do, or that the explanations can lead to novel predictions. In Part II of this book we will show that linear causal models can explain in the same fashion that, say, Daltonian atomism explained the law of definite proportions, or that Maxwell explained electromagnetic phenomena, or that Copernican theory explained regularities of planetary motion. We will also show that theories with latent variables can lead to novel predictions that would be unlikely to be found without them.

2.3.2. Interpreting Latent Variables

In the social sciences, latent variables are routinely given some gloss. They are labeled "socio-economic status," or "cognitive ability." In describing the TETRAD program, and in applying it to actual cases, we will follow the common practice, but with serious qualms. It is important to understand some difficulties with the practice, and to beware of certain confusions that may result from it.

Consider socio-economic status. In sociological economic status index will typically be the result of changes in all of the factors that enter into that index. The structural equations of a linear model typically apply to the values of the variables for *individuals* in the population (or sample), not to averages of values of the variables. (Many of these and other difficulties with the conventional use of socio-economic status as a latent variable have been sensibly discussed by Blalock , [9].)

For example, a familiar model for data from the summer Head Start program postulates an unmeasured socio-economic status as a common cause of father's education, mother's education, father's occupation and family income. Taken literally, the theory seems to make no sense, since the measured quantities must, if anything, *constitute* socio-economic status. The arrows go in the wrong direction. Suppose one abandons the interpretation of the latent variable in the Head Start model. Suppose one no longer regards that variable as an amalgam of education, income and other variables. *That does not mean that one should, or*

must, abandon the graph of causal relations postulated by the initial model. It means only that considerable caution, and more agnosticism, should be used in interpreting what that latent variable is. There is nothing amiss, logically or scientifically, in having latent variables that one is not quite sure how to interpret. Surely that is better than having latent variables with an interpretation that cannot be sustained. There is nothing wrong with supposing that something, for which we have no convenient name, but which is *not* socio-economic status, acts as a common cause of parental education, occupation and income.

The most important error that inappropriate interpretations of latent variables may occasion is the failure to consider alternative models. Often variables are clustered together, and specified to be the effects of a common latent variable, because there is some common content to their description. In a study of school children, for example, Maruyama and McGarvey [72] group together a number of variables because they have to do with popularity, and group other variables because they have to do with achievement, and others because they have to do with socio-economic status. Within the causal model, these groupings carry a causal significance. They say that these features have a common cause. That may be correct, but it is not necessarily correct. That schoolwork popularity and playground popularity are both forms of popularity does not imply that they have a common cause, "popularity", or that they have any common cause at all. There is no *a priori warrant*, therefore, for failing to search for other groupings of variables.

2.4. THE IMPORTANCE OF EXPERIMENT

Diana Baumrind, [3] says that "The objective of the traditional social psychology experiment is to enable the experimenter to infer unambiguously the existence and direction of causal relations by ruling out alternative causal explanations ([3], p.1290)." The objective may not often be literally achieved, but she is surely right that one of the advantages, perhaps the principal advantage, of experimental procedures is that they help to reduce the number of alternative causal models that might account for a body of phenomena. From a Bayesian perspective, experimental data may give narrower posterior distributions than non-experimental data (compare Leamer [59]).

The advantages of experimental control can be seen from simple combinatoric considerations, provided we suppose that the outcome of each experiment is unambiguous. If we have four variables of concern to us, then there are $4^6 = 9,216$ distinct possible causal graphs. If, however, we can control each variable experimentally, then we can consider them two at a time, making sure that other variables do not change, and test whether wiggling variable X makes variable Y wiggle, and whether wiggling Y makes X

wiggle. Twelve experiments, two for each pair of variables, will therefore determine the pairwise causal dependencies. Two experiments, which determine the dependency of X on Y and of Y on X, will reduce the number of alternative causal graphs for the four variables to $4^5 = 1,536$, reducing the initial indeterminacy by more than four fifths. Of course it will take a lot more experiments than that to determine the form of the dependency of X and Y over a range of values, and far more measurements still to determine all possible non additive relations among variables.

But experiments aren't always unambiguous, which is the point to which we next turn.

2.4.1. Zeisel on Experimentation and the TARP Experiment

In criticisms of causal modelling, judgements about issues of *principle* are often entangled with criticisms of particularly bad practices. In this subsection and the next, we consider two particular studies that illustrate what some critics find objectionable about causal modelling.

The Transitional Aid Research Project (TARP) was one of the largest social experiments ever conducted. The experiment, and the treatment of the data, are described in Rossi, et al., *Money, Work and Crime*, [82]. The experiment was intended to test the effects of financial support for newly released felons on recidivism rates. Payments to newly released felons in Texas and Georgia were made through the state unemployment commissions for a period of six months. Control groups in both states received no such payments. The outcome was nil: payments made no difference to recidivism. Rossi, et al. concluded that unemployment, which was a condition of receiving the payments in the experiment, increased recidivism, while the payments themselves decreased recidivism, and the two effects simply happened to cancel one another out. Hans Zeisel was originally on the advisory panel for the experiment, but resigned in protest over the treatment of the data. He objected, among other things, that there was a better account of the outcome, namely the obvious one that payments have no effect on recidivism.

Rossi et al. responded with various *ad hominum* arguments, [83], with the claim that they fully acknowledged and even emphasized the null effects of the payments as administered, and by citing a previous, smaller scale experiment which they claimed supported their interpretation. (That effort, the LIFE experiment, conducted in Maryland, is also discussed in their book). They further argue that their model is to be preferred to Zeisel's hypothesis because they thought of it first:

...the counterbalancing model and related specification were postulated before

the structural equations were estimated and were constructed both on the basis of the result of the prior LIFE experiment in which a payment plan did have a recidivism reducing effect and by drawing on social science theory concerning the possible competition between legitimate and illegitimate income producing activities. There is a vast difference between posing a theory before examining the data and posing one after examining them. On that basis alone, Zeisel's post hoc models cannot stand on equal footing with our a priori model. Furthermore, our modeling was tested through replication, being first constructed on the data derived from the Texas experiment and then tested on the data from the Georgia experiment.

On the substantive argument, Zeisel wins hands down. The LIFE experiment, performed under conditions that permitted those receiving payments also to seek legal employment, showed an 8% differential recidivism rate for crimes of theft, and no difference for other categories of crime. The difference was statistically significant, but only barely ([82], p. 37-43). The TARP experiment showed no significant difference. The only evidence the two experiments together provide for the effect of payments on recidivism is the weak evidence of the LIFE experiment, and that evidence is weakened further, not strengthened, by the TARP outcome. The temporal argument just cited is fallacious, as we have argued in a preceding chapter. It certainly makes an important difference if one theory uses the data to adjust a lot of parameters, and another theory accounts for the data without such fiddling, but in fact it is the hypothesis of Rossi, et al. that seems to require the most fiddling. But in this case, even the premises of the argument given by Rossi, et al., are false. The counterbalancing model was not obtained before examining the data. It is clear that the model was developed after the null effect of the payments was known, and was developed in order to save the experimenter's hypothesis in the face of apparently refuting evidence. The puzzling claim that the model was postulated before the structural equations were estimated (how *could one* estimate equations before obtaining them?) results from the authors' remarkable conception of scientific testing. Since they confuse estimation with testing, they believe there is some special virtue in having the equations before estimating.

The claim that the counterbalancing model was tested in any way is charitably described as Pickwickian. The conception of testing used by Rossi, et al. is certainly peculiar. They begin their section on "Testing the TARP Counterbalancing Model" ([82] pp.108-112) as follows:

So far the counterbalancing model...is a reasonable but not yet demonstrated hypothesis that can seemingly account for both the TARP and LIFE results. However, the model need not remain on the level of an unproven hypothesis since it is possible to use the TARP data to estimate coefficients for each of the links...

The authors then proceed to "test", "demonstrate" and "prove" their model *simply by estimating its coefficients!* No statistical test of the model is reported save for t tests of the coefficients. Any mathematically consistent, identifiable model whatsoever can pass the "test" of parameter identification, and the t tests of coefficients are only tests of the significance of the parameters *given* the assumptions of the whole "counterbalancing model."

The claim that the model is tested because the same model was used on both the Texas and the Georgia data is smoke in logical eyes. Any identifiable model can be estimated on two different samples. We are not told whether the small differences in coefficient estimates are significant, but sameness of the coefficient estimates in the two populations would show nothing more than that the measured correlations are the same. That might indicate that the data are reproducible, and give some confirmation of the appropriateness of linear modeling assumptions, but it has no other bearing whatsoever on the truth of the counterbalancing model.

Zeisel's methodological conclusion is more sweeping and less tenable than is his criticism of the arguments given by Rossi, et al. He claims that there is a difference in kind between correlational and experimental data, and that experiments are always unequivocal.

To interpret correlational data a theory is needed, and unless the theory is correct the interpretation will not be correct. It is the beauty of the controlled experiment that all the theorizing goes into its design. The result of the experiment, though more limited in its scope than correlational speculation, speaks for itself and needs no further theoretical support.

Would that it were true. It is not. It is not even close to the truth. Take any experimental subject you please, and its history and contemporary practice will show the reverse. Chemistry: Robert Boyle's "pigeon" experiment, in which water boiled in a closed flask left a solid residue, was taken by Boyle to show that water could be turned into earth. Others disagreed. Dumas' measurements of the vapor density of phosphorus and sulfur in the 1830s were taken by Dumas, and others, to refute the atomic theory. The experiment was interpreted consistently with atomism by Gaudin shortly after, but the interpretation remained controversial for thirty years. More recently, the interpretation of experiments that were purported to demonstrate the existence of a new physical state of water ("polywater") was controversial. Physics: In the late nineteenth century, the foundations of electrodynamics turned on the interpretation to be given to a large number of optical and electrical effects, effects that could be produced at any time in almost any laboratory or observatory. Not least among them was the outcome of the Michelson-Morley experiment, whose interpretation was doubly controversial: many physicists disputed what the

experiment demonstrated, and some (e.g., Miller) disputed that the experiment had revealed any real (null) effect at all. The significance of experimental comparisons of the solar spectrum with terrestrial spectra was disputed from 1914 into the 1920s, at least. Recently the interpretation of experiments that purported to detect gravitational waves has been disputed, and so has the interpretation of measurements of the shape of the sun. These are not isolated cases. They are typical cases.

Should one think that subjects, such as biology and psychology and sometimes even sociology, in which experimental and control groups are formed and given different treatments, somehow provide more certainty about the interpretation of their experimental outcomes? It seems unlikely that biologists, psychologists and experimental social scientists somehow provide a certainty that physicists and chemists cannot.

Control group experiments cannot be isolated from controversy for several reasons. Two prominent considerations are sample dissimilarity and unintended treatment effects. That samples are drawn from a population at random does not guarantee that they are representative. That subjects are assigned to various treatment classes at random does not guarantee that the class memberships are matched on all relevant non-treatment variables. Controversies can arise not simply because of statistical considerations about sample size, but also because of sampling methods, properties of the actual samples selected, etc. When a "treatment" is given to an experimental group, whatever is done is a complex event, with many facets. Separating out which facets are responsible for any differences found between experimental and control groups on outcome variables is not always trivial or obvious. The behavior of the experimenter's themselves may be the most important facet, as Rosenthal's [81] experiments suggest. The knowledge of the subjects themselves about their role in the experiment may be significant. It is exactly these concerns that are responsible for the double-blind design often used in medical research, in which neither experimenter nor subject knows which subjects are receiving the experimental treatment and which are controls. Even so, many medical experiments are anything but decisive.

The TARP study itself conflicts with Zeisel's opinion of the value of experiments. The explanation proposed by Rossi and his associates is *ad hoc*, and their arguments for it are egregious, but that does not show that their explanation is false. An *ad hoc* hypothesis can nonetheless be true, and bad arguments can be given for true hypotheses as easily as for false ones. Rossi, et al. in effect suppose that a lot of leisure and absence of work discipline increases the propensity of felons to commit new crimes, and that some cash in the pocket decreases that propensity, and the contrary effects are about equal. These are not absurd hypotheses, and they cannot be dismissed out of hand.

Zeisel and Baumrind seem to share the view that without experimental controls there is no science, and causal conclusions are never warranted. Their opinion is shared by others who do not bother to argue for it (e.g., Ehrenberg [22]). This is not a sustainable position. Most of astronomy developed without experimental evidence: Copernican theory was established without experimental evidence, as were Kepler's laws. Newtonian celestial mechanics had for *experimental* support only the law of the pendulum. The principal early argument for general relativity was its explanation of the anomalous advance of Mercury's perihelion, a phenomenon established by statistical analysis of non-experimental data using a lot of assumptions. The first clearly *experimental* evidence for the theory, Pound and Rebka's, was not obtained until 1960. Darwin had the experience of animal breeders, and his own uncontrolled experiments with pigeons, but his major evidence was the observation of variation and speciation in nature. Contemporary cosmologists observe. They cannot wiggle the entire universe, or build another cosmos for a control group. Good experiments are to be treasured, but science does not cease without them.

2.5. JUSTIFYING ASSUMPTIONS

2.5.1. Ling on Causal Modeling

Robert Ling's, [67], review of David Kenny's [56] *Correlation and Causation* is less a review than a call to arms against statistical and causal modeling. Despite its brevity, Ling's review seems to have caught the sentiments of the critics, and it is cited with approval both by Baumrind and by Freedman. The real substance of Ling's criticism turns on the existence of alternative models of data, and on the failure to search for such alternatives, a failure sometimes justified simply by appeal to "assumptions."

The logical fallacy underlying path analysis and other forms of inference from correlation can be illustrated by the following... A researcher believes that malaria may be caused directly by exposure to swampy air...Having specified the causal assumption by a path diagram, he finds a significant correlation between the incidence of malaria (Y) and the swampiness index of numerous locations sampled in the study. Ergo, the researcher concludes that "mal air" is the direct cause of malaria.

The foregoing example is not atypical of the manner in which theories are established by those employing the techniques described in this book. Not infrequently, the causal assumption (theory) is suggested by correlational data, which are then used (tautologically) as if the data were sufficient evidence to confirm the causal theory. In the path analysis methodology, the researcher can never disconfirm a false causal assumption, regardless of the sample size or evidence, so long as the variables alleged to be causally related are correlated.

There is a lot of hyperbole in these paragraphs. Of course the assumptions of path models and other linear causal models can often be tested, and often are. But testing is not nearly enough. Often the tests applied to causal models are of low power given the sample size, and that means that for all anyone knows there may exist many, many alternative models that would also pass whatever test has been applied. A set of assumptions in a causal model cannot be justified merely by statistical testing, without a search for alternative explanations of the data, or without direct arguments from other sources. It remains true, however, that some of the most objectionable procedures in the social science literature are exactly as Ling describes them. The TARP experiment is a vivid example.

The examples could easily be multiplied. They can be found in economics quite as well as in political science and criminology. Ling's accusations are a fair complaint in many cases. That doesn't make them good logical objections to the technique of causal modelling, or establish that path analysis is "at best a form of statistical fantasy." Failures of practice do not establish a failure of principle. Contrary to Ling, nothing in statistical causal modeling prevents researchers from disconfirming, or even rejecting, direct causal relations between correlated variables. From early days (see Simon, [91]) theorists knew that the hypothesis of direct causal relations could be disconfirmed by discovering other appropriate variables showing the association to be "spurious", i.e., the result of a common cause rather than a direct connection. Hypotheses about causal relations, direct or not, can be disconfirmed by the discovery of alternative models that give better accounts of the data, and do better by statistical tests. They can be disconfirmed by examining new samples from the population, or special sub-populations in which some variables do not vary, or by deriving from them predictions (or retrodictions) which prove to be false. Ling should have blamed some of the singers, not the song.

2.6. LINEAR THEORIES ARE LITERALLY FALSE

Anyone who thinks there is any point or justification for statistical modelling and for causal inference from non-experimental data should be prepared to give an example of good modelling and good inference. There are many dimensions of goodness, and a good study is not necessarily, or even usually, one that is uncontroversial or unobjectionable. It is one that enhances our predictive ability, or makes a persuasive case, or undermines our previous assumptions, and does so by reasonable scientific standards. By those criteria, it is easy to find good examples of causal inference from non-experimental data. In return, serious criticism of causal modelling methodology (rather than practice) should turn on the best cases, not the worst.

Blau and Duncan's [10] study of the American occupational structure was recently cited by the National Academy of Sciences [73] as an exemplary piece of social science. Since it makes heavy use of causal modelling techniques, it is the sort of example that critics should consider. David Freedman [28] has done just that. Freedman levels a great many criticisms, which we will consider one by one. His principal objections, however, are that Blau and Duncan do not justify the assumptions used in their model, and that the model is literally false. So let us first consider the value of literal truth.

2.6.1. The Aims of Linear Models

Nothing is more important in considering the value and limitation of a form of theorizing than keeping its *aims* clearly in mind. A procedure should not be dismissed for failing to do what it does not aim to do, or for not aiming and succeeding at what is impossible to do. In the case of linear causal modelling, the most important point is that little theories of this kind are *approximations*. They do not contain the exact truth or the whole truth.

Linear models are used throughout the sciences because they are conceptually simple, computationally tractable, and often empirically sufficiently adequate. They are almost never true in every detail. Of course a model that is not true in every detail can still be approximately true, and close enough to the truth to be relied upon in reasoning about action and about policy. What we want from social science is not truth in every detail, but theories that are close enough to the truth to be reliable guides in understanding and decision making. There need not be a unique theory that answers our need in any case. If two theories are both literally true, then they must be consistent with one another, but if two theories are each only approximately true, and thus literally false, they need not be consistent with one another. Any of several alternative theories may therefore sometimes meet our need for an approximately true theory on which to base action and policy. So long as each of them would lead to the same practical decision, it makes little difference which one we choose. But when the alternatives give different results relevant to decision making and action, it makes every difference which we choose.

This is not any kind of special pleading for weaker standards in the social sciences than in the natural sciences. In the natural sciences, nearly every exact, quantitative law ever proposed is known to be literally false. Kepler's laws are false, Ohm's law is false, Boyle's law is false, Maxwell's equations are false, and nearly every physicist believes that general relativity is false (because it is not a quantum theory), and on and on. These theories are still used in physics and in chemistry and in engineering, even though they are known to be false. They are used because although false, they are approximately correct. Approximation

is the soul of science.

2.6.2. Blau and Duncan's Model

Blau and Duncan's conclusions are often summarized with a simplified path model relating father's education, father's occupational status, son's education, status of son's first job, and son's occupational status in 1962:

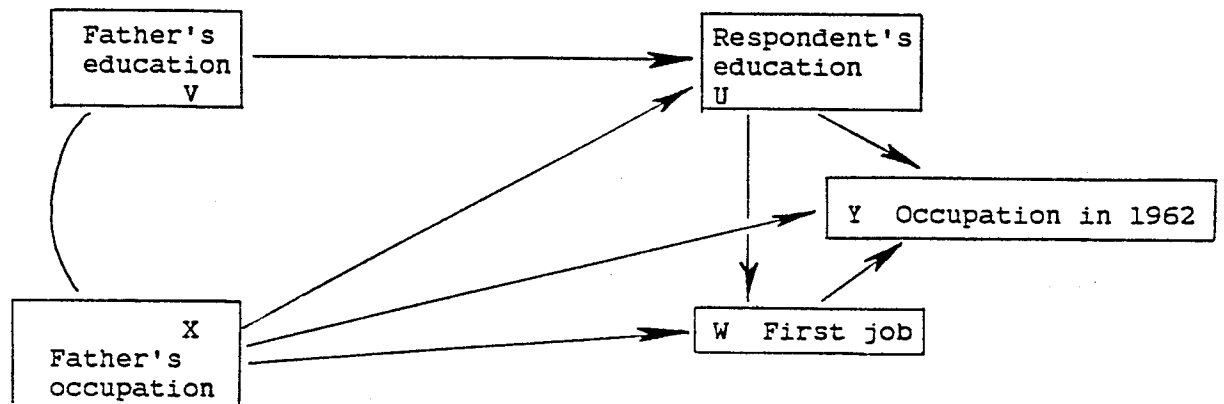


Figure 2-1: Blau and Duncan's Model

The model is elaborated in several ways in the course of Blau and Duncan's book.

Freedman observes, correctly, that Blau and Duncan do not explicitly give a system of structural equations and distribution assumptions. He charitably, and so far as we can see correctly, ascribes to them the set of structural equations obtained from the directed graph above by applying the rules we describe in Chapter 4. Freedman supposes the equations are meant to apply to individuals in the population, the error terms for different individuals are uncorrelated and normally distributed, and the error terms for distinct variables are also uncorrelated. Save possibly for the normality assumption, this seems a fair account of what Blau and Duncan intended. We will consider in sequence Freedman's objections, and some pertinent responses.

Objection 1:

There is no justification for the equations.

...the modelers have to make some showing that the structure of the equations reflects the structure of the phenomena...The only relevant

considerations are presented on pp. 166–168. In effect, Blau and Duncan argue that V_i and X_i are determined prior in time to U_i ; likewise, except "for an appreciable minority," U_i is determined prior to W_i ; and W_i is determined prior to Y_i . This is the whole of the argument..

Freedman suggests that the argument is supposed to lead to the structural equations by a recursive calculation of values of later variables from earlier variables, and he concludes:

Of course, this parallelism does not establish..[the equations]..as the right model, since many other systems of equations have the same recursive structure. For example, the effects could be quadratic rather than linear, or multiplicative rather than additive. To sum up, the equations proposed by Blau and Duncan do not have any adequate theoretical foundation.

Response:

Blau and Duncan have a pretty clear idea as to how both the structural equations and the stochastic assumptions are to be justified, and how justifiable they are:

As in the case with the assumption of linearity, we may for most assumptions adopt the pragmatic attitude that some departure from their literal truth may be tolerated if the assumptions facilitate an analytical objective. Yet it is often difficult to know if one has exceeded a legitimate level of tolerance and, especially, to comprehend what the consequences of sizeable violations of assumptions may be.

We have sought a way out of this dilemma that will put some burden upon the reader. Instead of using only one or two techniques, with attendant greater or lesser severity of assumptions, we have varied the techniques and consequently the assumptions. With some techniques we clearly go well beyond the point where the requisite assumptions can be at all rigorously justified. This venture, however, will—to the extent possible—be counterpoised by alternative treatments of the same data, avoiding at least some of the questionable assumptions. (pp. 116–117)

The response is, first, that the model is offered as an approximation. Indeed, the model Freedman focuses on is offered as a very first approximation, which is elaborated in the course of the book. Blau and Duncan never suggest that it is literally true. Second, they test their assumptions, linearity for example, against other, weaker assumptions that lead to nearly the same results. They explicitly address both the suggestion that the effects could be in some power of the variables and the suggestion that they could be multiplicative. The power function hypothesis is checked by performing a multiple classification analysis which assumes that the effects are additive but not necessarily linear (pp.132–139; for a detailed treatment of these and related techniques see [89]). It is further checked by

supposing that the effects are only piece-wise linear in categories of independent variables and showing that the slopes in the several pieces are very nearly the same, save for thinly populated extreme values of the variables. We find these arguments convincing. A weaker argument is given regarding multiplicative effects. Blau and Duncan argue that a logarithmic transformation of variables from similar data gave correlations very close to those they obtain without such a transformation (pp. 139- 140).

The strategy of using one set of assumptions, preferably weaker assumptions, to test other assumptions in a theory is ubiquitous in the sciences, and there is nothing wrong with it in principle, although like any strategy it can be practiced well or badly. It is the strategy Cannizaro used to argue for his system of atomic weights, and the strategy Jean Perrin used to determine Avogadro's number. Without it, much of the history of science would perish. The strategy, and some of its historical and contemporary applications, has been described in detail by Glymour , [34, 35].

Linearity is well short of entailing the particular theory Blau and Duncan propose with this model. They have a further argument, of which they perhaps do not make enough. No one, apparently not even Freedman, disputes that the variables they choose as exogenous and intervening have a direct or indirect effect on occupational status. The issue is what the causal relations are. The principal justification for the model is that it explains the correlation data for these variables very well, and *no alternative linear model seems available which gives a comparably good explanation of the correlations..* Even before the parameters are estimated, and independently of any normality assumptions, the model implies constraints on the covariance matrix that are very closely approximated in the data. After estimation, the model accounts for almost--but not quite--all of the empirical correlations.

Any alternative linear model must do as well. It must also satisfy the temporal constraints on the variables which Freedman alludes to. The real point of time order is not that it establishes Blau and Duncan's model, but that it constitutes a powerful constraint on any alternative. Freedman admits in a footnote that he could not find an alternative eight parameter model that fits the data as well as Blau and Duncan's initial model. Neither could we. Nor could we find a model with fewer parameters--such models tend to imply additional correlation constraints that are not very closely approximated by the data. A model with additional parameters can be made to fit the data, but such models either sacrifice simplicity while gaining nothing in their capacity to reproduce the correlations, or they sacrifice simplicity while losing the nice explanation of the constraints satisfied approximately by the sample correlations. Blau and Duncan's model is certainly not true, but it bears the marks of an elegant linear approximation to the truth, and that is even more

than they claim for it.

Objection 2:

The assumption that there is no autocorrelation in V or in X is violated by the sampling technique, since the data were collected from households, and the different respondents (all male) from the same household will surely have correlated values for these variables.

Response:

We cannot find any data in Blau and Duncan's book on what proportion of respondents were members of the same household, or how they were distributed. The objection is sound, but its importance depends on that information.

Objection 3:

There is no justification for assuming that the disturbance terms associated with the dependent variables are uncorrelated. There are, for example, famous dynasties in banking, politics and film, and the corresponding error terms for education, first job, and occupational status in 1962 will surely be correlated for members of such dynasties.

Objection 4:

Part of the error terms represent omitted variables, and these variables may act on several of the variables included in the model:

Still more generally, the model omits families, neighborhoods, geographical regions. It does not consider the quality of education, or when education was obtained, or when respondents entered the labor force...There is nothing of history in the model, and nothing of the economy.

Response to 3 and 4:

Of course one can find cases in which there is good reason to think there are common sources of variance for the dependent variables. Once again, the model is not assumed to apply universally and literally. Blau and Duncan explicitly consider in later models the effects of age, race, ethnic group, geographic region, farm versus non-farm background, marital status, father in law's occupational status, parental family size, birth order, family climate, and family type (divorced, not divorced, etc.), and more. Historical factors are discussed throughout the book, and historical considerations enter explicitly in the justification of the scales. It is true that these factors do not enter explicitly in the model Freedman attends

to, but why then complain of it when it is offered as a very first approximation, and the factors of concern are investigated as modifications and qualifications of the simple model?

Objection 5:

Significantly correlated errors will result in major changes in the estimated values of the parameters of the model.

Response:

Absolutely. Changes in specification will affect parameter estimates. There seems to be no evidence, however, that any of the variables Blau and Duncan consider produce a substantial correlation of the disturbance terms.

Objection 6:

Blau and Duncan make no predictions from their model.

Response:

No predictions are made. Perhaps some could be. Principally, that improvements in the educational level of a subpopulation in one generation will have little effect on the occupational status of their descendents if structural social factors prevent the members of the first generation from using their education to improve their occupational status, and if the educational improvements are not sustained in the second generation. It is, we suppose, one consideration pertinent to affirmative action and quota programs, although we have no idea what role, if any, Blau and Duncan's work has had on such practices.

Objection 7:

Their model ignores their own data analysis:

A fair summary is that the data clearly but narrowly violate the assumptions of the model: the regression curves are nonlinear (pp. 137, 144); the residuals are heteroscedastic (pp. 139, 144); the slopes vary across subgroups (p. 148). The path coefficients...therefore have no real existence. What are Blau and Duncan talking about?

Response:

Freedman's point is that Blau and Duncan's data analysis shows that their assumptions, in

particular their assumption of linearity, is literally false. It also shows that it is a *very good approximation* to the truth, and that is all Blau and Duncan claim for it. We should add that it appears to be at least as good an approximation to the data as are most of the simple classical chemical and physical laws. Physical chemistry would not be better off without the ideal gas law, or without the law of Dulong and Petit, but we very much doubt that the data available in the 19th century gave a substantially better fit for these laws than Blau and Duncan's data do for the linearity assumption.

Objection 8:

The model fails a bootstrap statistical experiment [21] of a correlation constraint it implies.

Response:

Of course the model fails a significance test. Any false model that is very close to the truth and implies overidentifying constraints will fail a significance test if the test is powerful enough and the sample size is large enough. In this case the sample size, 20,700 is very large indeed. The important thing is that any alternative explanatory linear model would almost certainly fare worse.

Objection 9:

Blau and Duncan claim that "the entire influence of father's education on son's occupational status is mediated by father's occupation and son's education", i.e., V has no direct effect on W or on Y . But the regression coefficient of V in a regression of Y on V , X , U and W is $-.014$ and the regression coefficient of V in a regression of W on V , X and U is $.026$, and both are statistically significant.

The conclusion drawn by Blau and Duncan is unwarranted...But a fair statement of their results is only as follows: Roughly, the data conform to the equations..., as depicted in the path diagram, although the differences are highly significant...Blau and Duncan seem to have been misled by their methodology into confusing assumptions with conclusions.

Response:

The objection that Blau and Duncan draw an unwarranted conclusion is pedantic. The regression coefficients are indeed *statistically* significant. They are also dinky, and in any ordinary sense they are not significant at all. Blau and Duncan's statement is a fair summary of the facts they find.

Objection 10:

The causal interpretation of path coefficients is nonsensical.

Suppose U , V and X are all functions of a more primitive variable τ , which is uniformly distributed over $[0,1]$. More specifically, let V be τ , let X be τ^2 , and U be τ^3 , but standardized to have mean 0 and variance 1...

Do we really want to say that the direct effect of τ on τ^3 is $-.611$, while the direct effect of τ^2 on τ^3 is 1.578 ? How can we vary τ while keeping τ^2 fixed?

The idea must be that structural equations are different from this artificial example. We need to have the difference spelled out.

Response:

It is essential to the usual representation of linear causal models that every variable have a unique exogenous source of variance. Freedman's "artificial example" violates this condition, and it is essential to his argument that it do so. That aside, suppose his conditions did actually obtain for some set of variables and some system. Then the causal claims would simply be false. That doesn't make the causal claims associated with real path models either false or nonsensical.

Real science has always proceeded by approximation and idealization. Many of Freedman's objections fail to appreciate that Blau and Duncan were doing, in a different setting, exactly what Newton and Dalton and Gay-Lussac and Hertz and Eddington did in theirs: approximating and idealizing, looking for simple, elegant, plausible, and probably not literally correct theories that explain the phenomena. By the standards the natural sciences impose on themselves, the *American Occupational Structure* is not bad work at all, not even statistically. One example: The Lick expedition of 1922 produced the best data on the gravitational deflection of light that were available until recent times. We guess that the gravitational effect accounted for less than 40% of the measured displacements of star images. The residuals of the least squares fit to the data were essentially unaltered if the relativistic deflection, which varies inversely with the square of the distance from the limb of the sun, was replaced by a simple inverse distance decay. Statistics does science many services, but it does no service at all if it keeps social science in thrall to a false and fantastic image of how science works.

2.7. CONCLUSION

The critical attack on the very idea of drawing causal conclusions from non-experimental data is miles short of convincing. Criticisms of actual practice are something else. The principal faults that beset linear modeling in practice, and that are seized on by critics, seem to be these:

1. Insufficient data analysis is done to justify the linear approximation and the stochastic assumptions.
2. Alternative models are not considered, and structural equations are not justified.
3. Inadequate consideration is given to variables that are not included in the model, but may be sources of covariance among variables included in the model.
4. No testable predictions are made.

TETRAD cannot help those who have not helped themselves about linearity and about stochastic matters. It can help substantially on the other counts. Later chapters demonstrate that the program can help to find good alternative models where they exist, and can help detect the existence of important neglected variables, although TETRAD will not tell the user what those variables are. We will show at least one way in which, using the program, testable, non-statistical predictions can be made from causal models of correlational data.

3. OBJECTIONS TO DISCOVERY BY COMPUTER

3.1. INTRODUCTION

The TETRAD program can be thought of as a device for peering into a vast space populated by causal models rather than by stars and planets. Some people would prefer not to look. There is a line of thought that objects to any computer program that claims to aid in the discovery of scientific theories. This chapter examines the arguments for that line of thought.⁷

Some objections apply to almost any form of computer discovery, or computer aided search for theories. The objections we will consider include the following:

- People have various kinds of special knowledge that computers do not have, and that knowledge makes people better at discovery than computers can possibly be.
- Discovery is a form of inference and inference should proceed in accordance with the requirements of Bayesian statistics and Bayesian epistemology, but computer programs that aid in scientific discovery often do not work on strict Bayesian principles.
- Offering computer programs that are intended to aid in the process of discovery is playing with fire. The programs are bound to be used stupidly and therefore should not be made available.

In addition, there are special objections to discovery procedures that are applied to statistical data. Such procedures must inevitably look at the data and use background knowledge together with structural criteria *and the data* to guide the search for the best explanations. Such procedures are routinely denounced as "ransacking" and "data mining," but *arguments* against the procedures are rarely given, and never given fully and explicitly. We identify four lines of argument, which we will discuss in detail in the third section of this chapter. In brief, they are:

⁷ Versions of these arguments have been given to us orally or in private correspondence by several people, and since our response is entirely critical and depends on reconstructing terse remarks, we avoid attributions save to published statements.

- There is a clear difference between "exploratory" procedures and "confirmatory" procedures. Exploratory procedures that examine the data to search for theories may provide theories worthy of investigation, but they do not provide any confirmation of those theories.
- It is circular or "tautologous" to use the data to search for the best explanation of the data, and then claim that the theory thus found is confirmed by that same data.
- Using the data to discover a theory contradicts the usual frequency interpretation of test statistics that are based on the same data.
- In the worst case, using the data to generate a theory will almost certainly lead to an erroneous theory.

The examination of this second set of arguments is considerably more technical.

3.2. THE GENERAL OBJECTIONS

3.2.1. People Know more than Computers Do

Objection:

Computers literally don't know what they are talking about. A computer procedure for searching for good theories must use structural criteria of some sort in its search, but we humans know a lot more than structure. We know what the variables mean, how they were obtained, whether some took their values prior to the time that others took theirs, and more. This knowledge is relevant to choosing a good statistical model, and a computer does not have it.

Comment:

There are two aspects to the objection. Because we humans know things about particular causal relations, or about the impossibility of particular causal relations, it may be thought that we can locate correct alternatives faster and more accurately than any computer can. But further, any computer program that uses structural criteria in its search, whether the criteria are statistical fit, or explanatory power, or whatever, will be likely to produce many absurd results, including causal hypotheses that we know cannot be correct.

Response:

The use of structural criteria without the kind of knowledge we humans have of causal possibilities and impossibilities can indeed be expected to lead to many absurd conclusions. The sensible thing to do is to combine human knowledge about a domain with the computer's structural criteria. There are two ways in which human knowledge can be combined with the computer's capacity to carry out systematic search. One way is to let the computer report the results of its search using structural criteria that are sufficiently inclusive to permit many alternative solutions to a problem. Then human users can employ their knowledge to narrow down the alternatives. The other way is to make it easy for humans to convey their knowledge about the domain to the computer, and then have the computer use this knowledge and structural considerations to guide its search. One of the major ideas behind contemporary expert systems is that knowledge specific to some subject matter can be programmed into the computer, so that the machine can use that knowledge in guiding its search.

The fact that without substantive knowledge structural criteria alone may sometimes lead to absurd hypotheses ought not to lead us to reject the use of those criteria when their results are consistent with what we think we know. Structural criteria encompass all of the usual virtues of theories, including simplicity, correct predictions, explanatory power, and so forth; in the absence of complete prior knowledge, we have no better means than these for forming preferences among our theories.

3.2.2. The Bayesian Objection**Objection:**

Artificial intelligence programs for discovering statistical models are really carrying out decision procedures, but they do not act like rational Bayesian agents. Even if a program did carry out Bayesian calculations, whatever prior probability distribution and utilities the program uses may not be shared by human researchers.

Response:

An artificial intelligence program might be designed to simulate a rational Bayesian agent. Programs of this sort have been developed for medical diagnosis (see Gorry, et al., [38]). But an artificial intelligence program might also be designed to help humans behave *more like* ideally rational agents. A program might well do the second sort of thing without carrying out any explicitly Bayesian calculations.

An ideal Bayesian rational agent is logically omniscient. The agent considers every possible hypothesis, assigns it a prior probability, assigns or determines a likelihood for the evidence on each hypothesis, and forms posterior probabilities by conditionalizing on the evidence. For many reasons, humans are not ideal Bayesian agents. As L.J. Savage, [85], noted, we humans are always falling short of coherence, and our struggle is always to get a little closer to it. In recent years, Bayesian statisticians have begun to consider explicit strategies for recovering from incoherence [65]).

One of the most important ways that we fall short of the ideal is in failing to consider alternative hypotheses. In applying Bayesian procedures to empirical data, we may form a collection (finite or infinite) of alternative hypotheses to account for the data, in such a way that each hypothesis determines a likelihood for the data. With proper priors, the sum or integral of our priors over all alternatives in the collection is equal to one. There is usually a further catch all hypothesis, namely that none of the hypotheses we have explicitly considered is correct. On reflection, we would rarely think that the catch all alternative has zero prior probability, but we do not know how to use it to assign a likelihood to the evidence. That is exactly because we generally do not know what the unexamined alternatives are, or what their mathematical properties may be. In the first chapter of this book we argued that this failure is one of the major limitations in sciences in which non-experimental or quasi experimental data are to be explained.

From a Bayesian perspective, the TETRAD program is a device for investigating a part of the catch all hypothesis, and for locating within it specific alternatives that give the evidence a reasonable likelihood, and that have the virtues of simplicity and explanatory power. These virtues can be thought of as utilities (see Kadane, [53], and Hempel, [44]), or as a constraint on prior probabilities (see Jeffreys, [49], and Zellner, [111]). Of course, a researcher might not share these utilities, or these priors, but we argue in a later chapter that they are fundamental desiderata in all sciences.

3.2.3. The "Fire" Objection

Objection:

Even if computer procedures for generating models are, if intelligently used, valuable aids to discovery, they should nonetheless not be made available exactly because they will not be used intelligently. Automated procedures may make it easier to fit arbitrary bodies of data, and to find models that appear to give good explanations for the data, even when the general modeling assumptions (linearity, for example, or absence of autocorrelation, or normality, etc.) are seriously in error. If so, they make it easier for people to obtain illusory

conclusions, and they make science worse.

Response:

The trouble with the argument is that it is generic, and it makes for terrible science policy. A perfectly analogous argument can be given against every technical innovation, computerized or not, that aids the enterprise of knowledge. There is nothing special in this regard about programs that aid in discovery. Regression packages make it easier for people to produce utterly inappropriate models when they neglect autocorrelation, multicollinearity, nonlinearity, etc. Estimation packages make it possible for people to assign numbers to ill-justified causal linkages, or even to confuse estimation with testing. Statistical tests make it easier for people to unjustifiably think they have established the truth of a model when it passes some statistical test of low power. Theoretical work, say Fisher's and Wishart's work on the distribution of correlations, or work on factor models, can have similar effects. Even programs for automated statistical diagnostics can make it easier to do stupid things. If one had a package that automatically checked for linearity and made appropriate transformations, and checked for autocorrelation, distribution assumptions, etc., it would make it all the easier for people to whip up good statistical analyses with stupidly chosen variable sets.

The same is true of technical innovations outside of statistics. Every physical instrument, from the telescope to the linear accelerator, has a variety of stringent conditions for its reliable use. If the user is unsophisticated, and the conditions for correct use are not met, a lot of garbage can and usually does result. Introductory physics students who repeat classical experiments with standard apparatus, Millikan's oil drop experiment for example, rarely get the accepted results. The history of science is riddled with hypotheses advanced because of inappropriate uses of new technologies. Should we therefore regret the introduction of the telescope, the microscope, the micrometer, the camera?

Every good technical innovation expands our capacities and presents new possibilities for discovery, new domains in which inquiry can be carried out. It also presents for the same reason the possibility of new errors, and new stupidities. There is absolutely nothing special in this regard about automated discovery procedures. The objection therefore rests on a more general policy, namely that the possible errors of the least competent members of a community are sufficient reason to suppress the introduction of new technical developments. Fairly, and therefore generally applied, that is a policy for ending science, not for furthering it.

3.3. NO PEEKING

3.3.1. Preliminary Remarks

Many social scientists and applied statisticians believe (or at least say) that one should never, or almost never, search through a body of data to discover theories. Instead, one should somehow come up with a causal model based on "theoretical considerations." Such people believe that computer programs that examine data and apply structural criteria and background knowledge to that data in conducting a heuristic search are objectionable no matter how well they perform in practice.

It is hard to make any reasonable case for this rather radical perspective. "Theoretical considerations," if they are to have any weight in guiding theory selection, must be well founded and therefore must be justified by some data or other. The position seems to be that only theories generated by not looking at any data, or only theories generated by not looking at any relevant data, can be true or can claim our serious attention. There is no rational reason to believe this sort of mysticism, however strongly it may be felt. Above all, the position denies what we know and what we can demonstrate: we know that people can consider only a tiny fraction of the causal models that are consistent with prior knowledge; we know that in the practice of statistical modelling, assumptions based on "theoretical considerations" are rarely well justified and that alternative explanations are rarely considered systematically; and we can (and will, in succeeding chapters) demonstrate that with computer aids, better theories can be found. Finally, we know that if close examination of data in order to search for explanations of it were prohibited, most of the natural sciences would not have developed. We would have to make do without the efforts of Copernicus, Kepler, Darwin, Cannizarro, and many, many others.

There is a more interesting and more precise kind of objection, not to computer search for theories, but to certain ways of conducting that search. This line of thought insists that while relevant data may indeed be examined in the process of searching for a theory to explain it, *the data used to discover a theory must be distinct from the data used to confirm or test or argue for the theory.*

The TETRAD program, like any discovery program, looks at data and searches for a good explanation of it. The causal models found by using TETRAD on a body of data can always be tested by comparing them with another sample, other than the one used in the search procedure. In the cases described in later chapters we sometimes do just that. So the program can always be used in such a way that the data used in discovery are distinct

from the data used in testing. But we also believe that *there is absolutely nothing wrong with using one and the same body of data to discover a theory and to confirm it or test it*. The arguments to the contrary are not inane or mystical, but they are erroneous or fallacious. We will consider four such arguments.

3.3.2. Exploratory and Confirmatory Procedures

Objection:

Computer searches are exploratory strategies and therefore cannot provide any confirmation of the hypotheses generated.

Response:

It has become routine in the social sciences to distinguish between "exploratory" data analysis procedures and "confirmatory" data analysis procedures, or between exploratory and confirmatory uses of data analysis procedures. The TETRAD program is perhaps most naturally classified as an exploratory procedure. We wish to point out, however, that the distinction carries with it a great deal of dubious intellectual baggage.

"Exploratory" is very often used to suggest that a procedure, or an application of a procedure, is useful for suggesting hypotheses to be tested, but that it does not of itself give any reason to believe any of the hypotheses suggested. Similarly, "confirmatory" is often used to suggest that a procedure does not suggest hypotheses, but that when somehow provided with a hypothesis, the procedure may provide a reason to believe it. Procedures that amount to parameter estimation and statistical hypothesis tests based on those estimates are called confirmatory, while other procedures usually are not. Except as a misleading terminology for distinguishing statistical hypothesis tests from other procedures for drawing conclusions from data and background information, this distinction is illusory. If an "exploratory" procedure routinely turns up hypotheses that do well by statistical tests and make accurate predictions, and if the procedure rarely turns up hypotheses that do poorly by such criteria, then the fact that a particular hypothesis is turned up by the procedure provides some substantial reason to believe the hypothesis, or at least to give it more credence than those hypotheses that are rejected by the exploratory procedure.

The notion that "confirmatory" procedures such as statistical hypotheses tests provide some substantial reason to believe hypotheses *independently of exploratory procedures* is mistaken, at least in the case of linear models. On small sample sizes tests such as chi square have low power against the class of all alternatives. Unless all but a handful of

alternative hypotheses have been ruled out, the fact that a model passes a statistical test provides almost no reason to give it credence, for there may well be billions of alternative models of the same data that, were they to be subjected to statistical testing, would do as well as or better than the particular model under consideration (see Chapter 1). We cannot rationally pretend that the alternatives don't exist, or that they are inconceivable, or unconceived. If we have no knowledge as to whether or not better theories may be hiding among the unexamined multitudes, how can it be reasonable to give credence to the theory we happened to think of and to examine statistically? But when the sample size is large, and the test has high power, the linear model is usually rejected, exactly because it is at best an approximation. Thus Blau and Duncan's model of the American Occupational Structure, which accounts for almost all of the correlations among its variables, is applied to a sample size in excess of 20,000, and fails various statistical tests (see Fox, [26], Freedman, [27]). This means that the p values of chi square statistics for linear models can only be used *comparatively*, but to do that we must discover the alternatives for comparison.

In fact, of course, we haven't either the human time or the computer time to do statistical analyses on more than a small number of alternative models. That fact makes it all the more important that statistical tests, if they are to be used, be used in conjunction with exploratory procedures that can search heuristically through enormously large numbers of possible models to discover those that will provide the best explanation of the data.

3.3.3. Equivocating over the Model

Objection:

A procedure that searches for a good theory, using the data that the theory is supposed to explain, and then argues that the theory deserves credence because it explains that same data, is tautologous.⁸

Comment:

We believe that the principal intuition behind the objection is that in order for a model to be confirmed it must be tested, and in order for a model to be tested something must be done that *could have* disconfirmed the model. Suppose a model is first conjectured and then the data is obtained to test it. Then even if the data actually confirms the model, if the data *had* turned out differently it *would have* disconfirmed the model, and so

⁸ See, for example, Robert Ling's review of Kenny's Correlation and Causality, [67]

something has been done that *could have* provided disconfirmation. In contrast, if a model is generated from a data set, and then "tested" on that same data set, nothing has been done that could have disconfirmed the model.

Response:

If the argument were sound, everyday counting procedures would be invalid. Imagine that the task is to determine the number of people in a room. We do that by counting, a procedure that involves generating hypotheses algorithmically as the data changes. We don't think of the procedure as "tautologous", and we are right not to do so, even though different data would cause us to generate different hypotheses. It looks as though something is wrong with the argument.

The objection depends on an elementary logical mistake: equivocation. Let's suppose that an ideal search procedure would use data but would never yield a model in conflict with that data. Now clearly distinguish two different claims:

- (I) No possible data, when given to an ideal search procedure, will result in a model that is disconfirmed by that same data.
- (II) No possible data can disconfirm the model produced by an ideal search procedure that uses some particular body of data.

I is true, but II is false. The objection confuses the true claim, I, with the false claim, II, and the argument turns on that confusion. The mistake is to confuse the true claim that the generation procedure cannot possibly yield a model that is disconfirmed by the data used to find that model, and the false claim that nothing has been done that could have disconfirmed the model actually generated. If variables are measured and a model, call it *M*, is generated from the data obtained, something *has been done* that could have disconfirmed *M*: the measurements could have turned out differently and if they had turned out differently they would have disconfirmed *M*. The situation is in all relevant respects exactly like the situation in which the model tested is generated without examining the data. It is true that nothing has been done that could result in *the generation of a model* that is disconfirmed by the data (at least not if the generation procedure is an ideal one). But that is not the issue; the issue is whether anything has been done that could have disconfirmed the model actually generated, in this case *M*, and something of that kind has been done. If the data had turned out differently, *M* would not have been generated, exactly because *M*

would have been disconfirmed. The objection turns on failing to keep track of what is being talked about.

Consider counting again. If there are twelve people in a room, the counting procedure will generate the hypothesis that there twelve people in the room, and will stop with that hypothesis. The data will confirm the conclusion. If there had been only ten people in the room, the hypothesis that there are twelve would not have been generated by the counting procedure, exactly because it would have been in conflict with the facts.

3.3.4. Difficulties with Frequency Interpretations

Objection:

If the model is found using the data, and the model is estimated and tested using that same data, then the standard errors of coefficient estimates and the p values of statistical tests may lose their usual meaning.⁹

Comment:

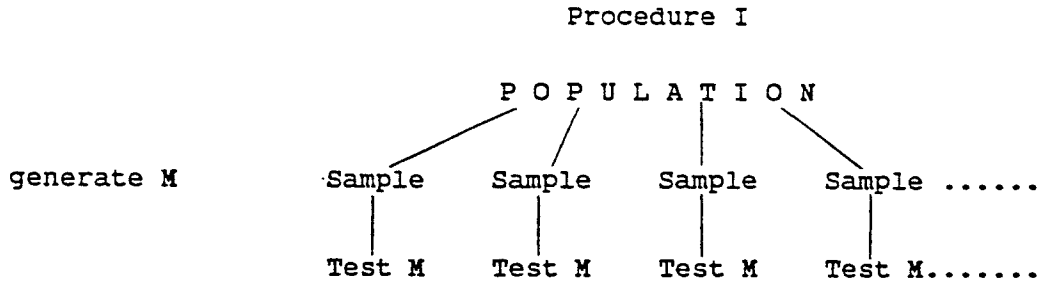
Suppose a coefficient in a linear model is estimated by a maximum likelihood estimator. Given the sample size and the assumption that the model correctly describes an infinite population, the estimator has certain *mathematical* properties, such as its variance. Similarly, given a model, data, and a sample size, the chi square statistic for that model, data and sample size has certain mathematical properties, such as its p value. The clearest and least controversial meaning of these statistics is that they are simply mathematical relations between the stochastic model and the data set.

Frequentists give these mathematical relations a further interpretation, and it is this interpretation that is applied in the objection. Suppose we have a stochastic model, M , a data set of size n randomly selected from some population, and a probability value, p , for a chi square statistic calculated from the model and the data. The interpretation is that the p value of the statistic is to be understood as the long run frequency (or limiting relative frequency) with which a value that extreme would be obtained in a sequence of random samples of the same size, n , drawn from a population that is truly described by the model M . An analogous long run frequency or limiting frequency interpretation is given to the standard errors of coefficient estimates.

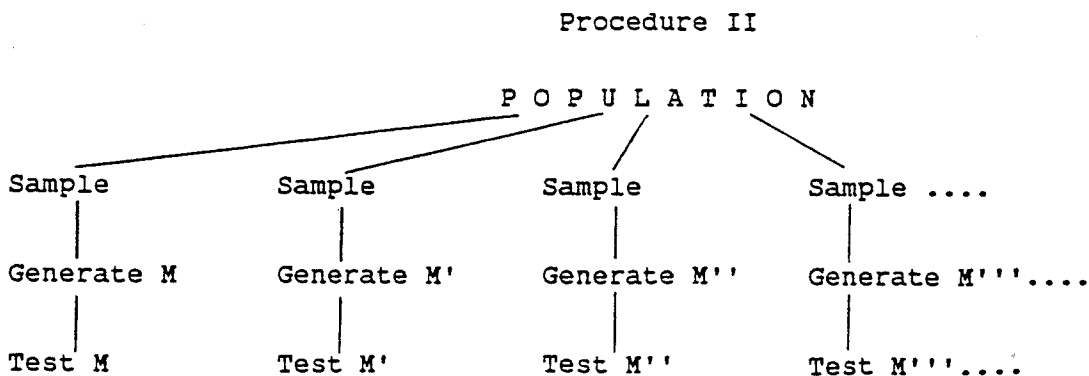
⁹Bentler, [6] and Bentler and Lee, [60]), make this objection to the practice of standardizing models on the same data that is used to estimate coefficients or perform chi square tests of the models.

The point of the objection is just that if the procedure we use involves somehow estimating the model from the data, then the p values of tests and the variances of estimators *cannot* be given this frequency interpretation. For if we imagine a sequence of samples drawn from the population in which each sample is used to discover a model and calculate a chi square value and estimate coefficients, then *the model found will typically vary from sample to sample in the sequence*. At each step in the sequence we won't be estimating the same coefficients as occur in M, and we won't be determining the value of the chi square statistic for one and the same model M.

The objection can be put pictorially. The frequency interpretation of the p value of a test statistic for M for a given data set is the frequency with which we would get a more extreme value for the statistic in a sequence like this:



But if the model M we generate depends on the sample data we obtain, then the procedure corresponds to the sequence:



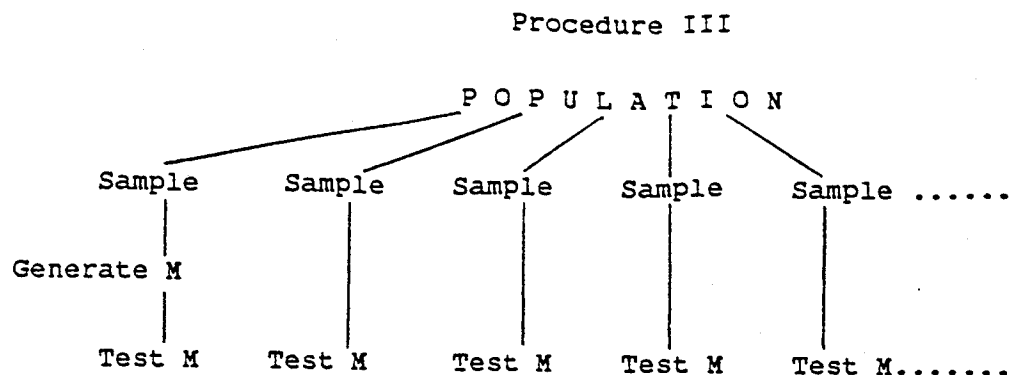
The p value of the chi square statistic as conventionally calculated does not in general have anything to do with the long run frequency of values of chi square statistics in a sequence of this second kind.

Response.

The argument is a complicated muddle. First, note that except in Monte Carlo studies and the like, sequences of the sort described are almost never actually carried out. We do not do large numbers of repeated samplings of Head Start students, for example, and apply one and the same model to them, or a model generating technique to them. In practice we look at no more than a handful of data sets for any collection of variables. We may, or may not, change the model we apply to later data sets based on how our initial models perform when tested on earlier data sets. We do not choose a model, draw a sample from a population, standardize the model and estimate it, choose another sample, and so on infinitely, or even for a long while. In practice we argue over alternative explanations for a few data sets, and move with our statistical or algorithmic techniques from problem to problem.

So the attachment of long run frequencies to p values and standard errors is not a *description* of scientific practice. Instead it is a way of trying to *interpret* what the statistics *mean*. If you think that probabilities are mysterious but long run frequencies are not, then assigning a long run frequency to a probability claim is a way of making sense of the probability.

We don't wish to challenge the frequency interpretation of probability. But we claim it applies just as well to the p values and standard errors of models obtained by looking at the data as to models obtained by not looking at the data. We claim that for a particular model, M, obtained by examining data set D, the p value of a statistic calculated from M and D *can* be given a frequency interpretation just like that in the first sequence shown above. Suppose you draw a random sample D and use it to find a model M, and calculate a value for the chi square statistic. Now the p value can be interpreted as the frequency with which a value as large or larger would be obtained in a sequence just like the first one above, except for the first trial in the sequence:



So, if you want a frequency interpretation of the p value calculated for a model generated by looking at the data, there it is, as clearly as it is for models generated without looking at the data. Moreover, so far as the frequency interpretation of probability is concerned, Procedure I and Procedure III are indistinguishable.

Those who make the objection under discussion will very likely fuss at this point, and say something like the following:

Further Objection:

Every stage in the sequence associated with the frequency interpretation of a statistic obtained by a procedure must reproduce all of that procedure. If the procedure generates a model from the data, then every stage in the sequence associated with the statistic must also generate a model from the data. So Procedure II must be used, not Procedure III, and Procedure II does not give a frequency interpretation of the p values or standard errors of a fixed model, M.

Further Response:

The appropriate response is quite simple: why? A frequency interpretation is wanted, and one has been given in Procedure III. Moreover, it is an interpretation in which the frequency agrees with the mathematical p value. Why insist that the only appropriate frequency interpretation is the *wrong one*, namely the second of the three sequences we have shown? We think the answer to this question is that two quite different concerns are behind the objection, and they have been confused.

One concern is to give a frequency interpretation of the statistics. We have shown that to be a trivial matter. The second concern is with the *reliability of the procedure* and that is

a quite different and much more complicated matter. Under the Neyman–Pearson account of hypothesis testing, we may choose a significance level such as .05, and decide to reject a model M if it yields a test statistic with a p value smaller than that significance level. The significance level tells us, under the frequency interpretation, the frequency with which we would erroneously reject M in a sequence of tests using samples drawn at random from a population correctly described by M . The significance level tells us something about the reliability of the testing procedure. It tells the probability that we will reject M , given that M is in fact true. The procedure for generating the model has nothing to do with this decision theoretic account of testing the model, and it can be applied to the testing of computer generated models just as the frequentist interpretation can be applied to the statistics for such models. It applies to the model M in Procedure III as well as to the model M in Procedure I.

But we can quite appropriately wonder about more than the reliability of a testing procedure once a model is generated. We can instead ask about the reliability of any procedure for selecting theories, where the selection process may consist both of generating models and testing them. Suppose our Procedure II uses sample data and background knowledge to generate a model (or a set of alternative models) and then computes the value of some statistic for each model and rejects models whose probability value for the statistic is too small. What is the probability that the true model will not be generated, or will be generated but then rejected? What is the probability that a false model will be generated and not rejected? The p values and significance levels don't tell us the answers to these questions, because they don't tell us anything about the probability that a procedure generates the true model.

All of this is perfectly true. *But it is as true for humans as for computers.* Imagine a statistician who steadfastly refuses to look at the data when he is constructing a model to explain it. He or she generates a model then looks at the data and tests the model. We can think of the statistician as a biological computer that generates hypotheses and tests them. What is the probability that the statistician will generate but reject the true hypothesis? That information is not given by the significance level of the test statistic applied. In practice humans use a complicated and unknown procedure for generating models. They use information about prior samples to choose the models they will apply to new samples, they use their prejudices and their imagination, and, often enough, they even look at the data the model is to explain.

The appeal of the objection lies in the almost automatic assumption that there is something special about hypothesis generation done by people. But human beings go through some sort of complicated procedure, the details of which we do not know, in selecting a model

to be tested. Presumably one explicitly rejects some models, fails to think of other models and therefore implicitly rejects them, and so forth. If we perform a chi square test of the model generated by a human, we have of course not determined much about the probability that the human failed to consider, or prior to testing considered but explicitly rejected, the true theory. Whether the theory is generated mysteriously by a person, or less mysteriously by a computer, the significance levels of test statistics mean the same thing, and they do not mean the probability that the true theory has been rejected. They mean, on the conventional interpretation, the probability of rejecting the null hypothesis, on the assumption that it is true.

Leaving aside the specific Neyman–Pearson framework in which probabilities are frequencies, it makes sense to ask about the probability that a hypothesis generation procedure will generate the true model for a sample of fixed size drawn from a given population. If a model generation procedure is algorithmic, then there should be a mathematical fact of the matter as to the limiting frequency with which true models are generated in a sequence of the second kind (Procedure II). We can expect such probabilities to be hard to calculate but one could approximate them with a long but finite sequence of Monte Carlo trials. But in the same way, we can investigate the performance of any human model generator in a long but finite sequence of the second kind. And, we can likewise investigate the performance of a human aided by a computer search. *The essential practical question is not the absolute probability that an algorithmic procedure will generate a true model, but the comparison of the algorithmic procedure's behavior with the behavior of unaided humans and with computer aided humans.* We have no experimental evidence to offer on this question, but we do offer in later chapters a great deal of indirect evidence that people using the TETRAD program do better than people without it.

3.3.5. The Worst Case Objection

The objections we have considered so far rest on common misunderstandings. But there is an objection to computer aided construction of statistical models that does not depend on any misunderstanding. We will give it first informally, and then more carefully.

Objection:

An algorithmic procedure for finding models by examining the data and applying heuristic search will produce *some* model, even for data that are in fact randomly generated from independent variables. With enough variables and a small enough sample, correlations will appear just by chance, even though the variables are independently distributed, and given

such data an algorithmic procedure will find a causal model that accounts for the correlations, even though they are in fact due to chance. Algorithmic procedures guarantee that a model will be found that appears to account for the data, even if there is no true account to be given beyond the chance effects of the sampling procedure.

Comment:

A little more precisely, the argument seems to have the following form:

1. Computer aided heuristic searches for statistical models must examine the data for statistical dependencies among the variables, search for the model or models that best explain those dependencies, subject the models thus obtained to statistical tests based on the data, and output those models that survive the tests.
2. No procedure for searching for hypotheses is acceptable if there are circumstances in which it is very probable that that procedure will yield a false conclusion.
3. For any procedure as in 1., a number r of independent random variables and a sample size n can be found such that it is very probable that a sample of size n will show k statistically significant correlations (or other statistic) among h of the r variables, for some number h and for some number k . That is, it is highly probable that at least k correlations will be sufficiently high that the probability of correlations as large as *those* occurring by chance in a sample of size n drawn from a population of values of the h random variables is less than .05, or whatever significance level is chosen.
4. In the circumstance described in 3., it is very probable that a procedure such as is described in 1. will output false hypotheses.
5. Therefore, by 4. and 2., a computer aided heuristic search procedure is unacceptable.¹⁰

Response:

This argument reveals a real problem about discovery. Before explaining why we think the argument should nonetheless be rejected, we wish to point out that even if one is completely convinced by the argument, it provides no reason not to use the TETRAD program. For TETRAD, part of premise 1 of the argument is false. The argument depends on the assumption that the models generated by heuristic search will be subjected to a statistical test using the very same data that were used by the program in its search. Nothing requires that TETRAD be used this way. The program can as easily be used to

¹⁰ This argument is modeled on a similar argument about estimation given by Ronald Giere [32].

generate models using one sample, and the models it generates can then be tested statistically on another sample. We have used this procedure, for example, in our study of the summer Head Start data, described in a later chapter. Anyone convinced by the preceding argument can use TETRAD in this way. But no one ought to be convinced.

Premise 2. is incorrect, for two reasons. First, it puts all of the weight in judging a procedure on the desirability of avoiding *false* theories. We think that is desirable, to be sure, but it is at least equally important in judging a procedure to consider the desirability of finding *true* theories, (see Levi, [61]), or approximately true theories.

If the only criterion for comparing methods is that of avoiding commitment to a false theory, then the optimal strategy is clear: accept no theories, no matter whether they are discovered by humans, by humans with computer assistance, or by computers alone. Instead of any such strategy, we prefer to weigh the desirability of avoiding commitment to a false theory against the desirability of not overlooking commitment to true theories, or approximately true theories. Computer programs such as TETRAD help us to avoid overlooking theories that may turn out to be correct. They therefore assist us in avoiding an important kind of error, and that contribution should not be ignored in assessing the advantages and disadvantages of computer generated theory.

Second, premise 2. assumes that a procedure ought to be judged by the *worst imaginable case*. We think procedures should be judged by the *expected case*. In the majority of cases researchers are pretty confident that the statistical dependencies they find are due to some causal structure or other, even while they may be much less confident about any particular explanation of the data. In the vast majority of cases, if the investigator were not strongly inclined to think that there is some explanation other than chance (or bad measurement design) for the patterns found in the data, a causal model would not be sought in the first place. Unless the researcher thinks there is a large probability that the dependencies in the data are spurious, there is no sufficient reason not to use the data to search for the best explanation of it. Of course, some of the correlations found may be due to chance, and that is the more likely the smaller the sample size in proportion to the number of variables considered. The investigator should certainly take account of that fact, and where appropriate test a model on new samples.

These theoretical arguments overlook the reality of practice. In practice, many social and behavioral scientists behave much like natural scientists. Studies of really important questions are repeated, and the dependencies that demand explanation are the robust ones. In practice, researchers in the social and behavioral sciences, like their colleagues in the natural sciences, will inevitably look at the patterns in the data in search for the best

explanation of it. Inevitably, alternative explanations of the same data will be offered by a series of workers, provided the questions at stake are important enough. That is the procedure in every empirical science worthy of the name. Artificial intelligence procedures only aid people in doing better what they will do anyway, and, in most cases, ought to do.

In practice, TETRAD does not make the user overconfident. We find that people who develop models without the systematic heuristic search provided by the computer typically place too much confidence in particular models. The computer does not mislead us by generating a single, false theory, when in fact we should withhold acceptance from all theories. Instead, the computer often warns us, when we have data that do not warrant any particular conclusion, that there are a great many different ways to explain the phenomena, none especially better than the rest, and that we are therefore not warranted in accepting any particular conclusion.

Even while we regard the last argument given against heuristic search as unsound, it does provide an important caution. Whether in astronomy or sociology, some statistical dependencies will occur by chance, and if we account for them by causal processes, then we will make a mistake. That is part of the burden of science: sometimes we are going to be wrong.¹¹

3.3.6. Explanation and Prediction

The strong conviction that "exploration" or "search" must be separated from confirmation or testing has a long history, but the arguments for the separation are seldom clearly formed. The conviction is almost universal in applied statistics, shared by Bayesians and more orthodox statisticians alike. Ultimately, the desire to separate the data used in discovery and in justification, in theory generation and in theory testing, probably derives from the conviction that only correct *prediction* counts as confirming a hypothesis. Explanation after the fact is discounted, on the grounds that it is, or may be, *ad hoc*.

Something of the same prejudice is occasionally found in the natural sciences, but it is much less prevalent. The DENDRAL program [66] examines mass spectrograph data to find the best explanation for that data, and chemists do not find this practice objectionable. Physicists routinely reexamine old data in search of new explanations. Recently, for example, data from one of the classic tests of the general theory of relativity, the Eotvos experiment, was reexamined to argue that the principle of equivalence, which is fundamental to the theory, is false. Einstein himself argued for the general theory on the grounds that

¹¹The technical worst case argument against procedures that generate theories by examining the data can be defeated by taking a sufficiently large sample size. Thus in the case of Blau and Duncan's study of the American occupational structure, with a sample size in excess of 20,000, there is virtually no chance that the correlations and other dependencies among half a dozen variables are due to chance.

it explained data about irregularities in the motion of Mercury, irregularities that had been known for sixty years prior to the publication of the general theory in 1915, and that had been the subject of many alternative explanations. In fact it is clear that Einstein used the anomalous motion of Mercury and the equivalence principle as guides in his long search for a satisfactory relativistic theory of gravitation. Copernican astronomy was founded almost entirely on historical observations of the sun, moon and planets. Most of these observations had been made centuries and even millenia before Copernicus wrote, and had been used to obtain other theories, notably the Ptolemaic theory and its modifications. Cannizzarro's powerful argument for the (then controversial) atomic theory in 1860 was not based on any newly confirmed predictions, but on a systematic review of the evidence that had been accumulated in the preceding decades, and on the argument that the atomic theory provided the best explanation for patterns revealed by that data.

Whatever the rhetoric, most physical scientists act as if a good explanation can be nearly as valuable as a good prediction, and they know that most of their science would not have emerged if their predecessors had been forbidden to examine the data in searching for theories, or if data once used to test a theory were thereafter forever tainted and useless for confirmation.

People rightly worry that a theory constructed by someone after seeing the data will be constructed specifically to account for that data. If the theory is deliberately constructed to account for the data, then the data provides no test of the theory, and hence no reason to believe the theory. This worry is well founded, but misstated. The notion of "constructing a theory to account for the data" is complex. In one sense people certainly do construct theories; they make them up, piece them together, put them forward. In another sense, a theory is simply an abstract object, like sets or numbers. The abstract objects may be discovered, but discovering them isn't creating them. theories are out there and have whatever logical relations they have with the data, whether or not anyone happens to think of them. Either a theory has the right logical relations with the actual evidence, in which case the evidence confirms it, or the theory has the wrong logical relations with the actual evidence, in which case the evidence disconfirms it. How a particular human being happened to discover the particular theory has nothing to do with the matter. The job of a computer program for discovering theories is to find the theories with the right logical relations to the evidence.

How can this understanding of confirmation and testing be reconciled with the powerful sense that some theories are unsatisfactory because they are "cooked" to account for the data? The answer is that the "right" logical relations to the evidence consist in a lot more than merely being consistent with the evidence, or entailing the evidence. Cooked theories

may entail the evidence but they don't entail the evidence in the right way. Put slightly differently, if the evidence could have been predicted using one theory, but not using another, then the first theory typically provides a better explanation of the evidence.

The succeeding chapters describe a conception of explanation that applies to linear causal models, and that is more robust and demanding than the requirement that the theory fit the observed statistical dependencies. The TETRAD program helps to search for theories that provide such strong explanations, and while more evidence is always better than less, we believe it makes no difference of principle whether the theories thus found are tested statistically on the same data or on new data.

PART II

The TETRAD Program

4. CAUSAL AND STATISTICAL MODELS

4.1. INTRODUCTION

Our principal concern is with procedures for *specifying* a model or models to account for non-experimental or quasi-experimental data. We are especially interested in procedures for specifying the causal hypotheses implicit in a model, for it is these hypotheses that are often of chief importance in practice, and reliable statistical estimation usually depends on having the correct causal structure.

Most linear models in the social and behavioral sciences have a causal interpretation, and often the causal claims within a statistical model are the principal concern and point of the investigation. Sometimes, of course, no causal interpretation is given to a model, or is appropriate. More often, the causal claims in a model are signaled by special phrases. Psychologists often mark causal relations by specifying that some variables are *dependent* and others are *independent*. Econometricians, and increasingly other social scientists, mark causal relations by specifying that some variables are *endogenous* and others are *exogenous*.

Linear causal models include a set of equations relating the variables of the model, a set of stochastic assumptions about the probability distributions of those variables, jointly and individually, and a set of causal relationships among the variables. Factor analytic models are of this sort, and so are path analysis models, structural equation models, models with random variable coefficients, and many regression models and econometric models. *The strategy behind the TETRAD program is to abstract the causal structure from a statistical model, ignoring the equations and most of the statistical assumptions.* We call this abstract structure a **causal model** to distinguish it from the fuller statistical model containing it. The causal model consists of a **directed graph**.

It is possible to work with a very simple mathematical structure, the directed graph, rather than with the more complex equations and distribution assumptions of a linear model, because the directed graph of hypothetical causal relations determines important statistical properties of any linear model that contains it. The theoretical basis of the TETRAD program consists of a number of new theorems about the connection between directed graphs and statistical properties of models. These theorems are described in this chapter and proved in chapter 10 of this book. Because of these connections, in later chapters we will be able to ignore most of the usual mathematical paraphernalia of linear models and

focus on graph structures.

The aim of this chapter is to explain how causal hypotheses can be represented by directed graphs, to describe the relationship between directed graphs and the usual forms of structural equation models, and to describe some of the most important statistical properties that are determined by the directed graph of causal hypotheses in a model.

4.2. SOME TERMINOLOGY

Linear causal models usually contain parameters whose values are not specified. For example, a model may contain an equation that specifies that variable X is a linear function of variable Y, but the linear coefficient may not have a specified value. Such free parameters must be estimated from the correlation or covariance data of samples drawn from the population to which the model is supposed to apply. When we write of a **statistical model** we mean a model that may contain free parameters. These parameters can either be constants with an unknown value, or random variables. When we use the phrase **estimated model** we mean the result of specifying the values of all free parameters in a statistical model, using appropriate sample data. Statistical models with free parameters do not entail actual (non-zero) numerical values for the correlations or covariances of their variables, but they can entail that the population correlations or covariances must satisfy certain equations. These equations are sometimes called **overidentifying constraints**. Estimated models imply definite numerical values for the correlations or covariances of their variables, and these values can be compared with the correlations or covariances exhibited by samples drawn from the population.

4.3. DIRECTED GRAPHS AND CAUSAL MODELS

4.3.1. Directed Graphs

Hypothetical causal relationships implicit in a model may be represented as a **directed graph**. A directed graph is simply a list of pairs of variables. The first member of a pair is assumed to be a direct cause of the second member of that pair.¹² Directed graphs can also be represented by line drawings. For six variables connected with socio-economic status, say socio-economic status (SS), father's education (fe), mother's education (me), father's occupation (fo), mother's occupation (mo) and family income (in), one simple causal

¹²Variable X has a "direct" effect on variable Y in this sense provided the variation in Y in the population is due in part to a variation in X that is not mediated by any other variables that are explicit in the model.

theory can be represented by a list:

$$\{ \langle SS, fe \rangle, \langle SS, me \rangle, \langle SS, in \rangle, \langle SS, fo \rangle, \langle SS, mo \rangle, \langle fe, fo \rangle, \langle e1, fe \rangle, \langle e2, fo \rangle, \langle e3, me \rangle, \langle e4, mo \rangle, \langle e5, in \rangle \}$$

or by a picture:

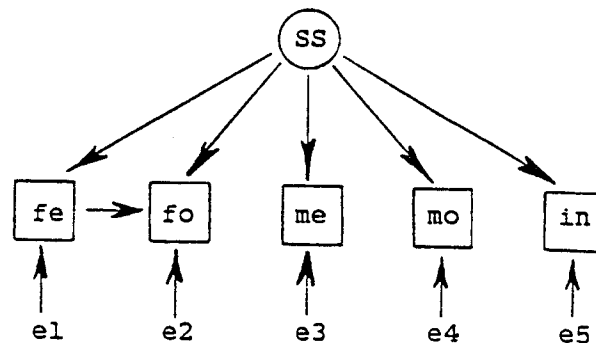


Figure 4-1: Directed Graph Representation

An ordered pair of variables that is in the list of a directed graph, or that is connected by an arrow in the drawing, is called a **directed edge**. We can think of a directed edge pictorially as simply the arrow connecting two variables. The variables connected by the directed edges are **vertices** of the directed graph. A directed edge is said to be **into** its second vertex (the head of the arrow) and **out of** its first vertex (the tail of the arrow). Two vertices connected by a directed edge are said to be **adjacent**. The number of edges directed into a vertex is its **indegree**; the number of edges directed out of a vertex is its **outdegree**.

4.3.2. Paths and Treks

Sequences of directed edges that connect variables specify a hypothetical path of causal influence from one variable to the other. More exactly, we define a **path** from variable u to variable v as a sequence of directed edges $u \rightarrow w \rightarrow \dots \rightarrow v$, with all arrows running in the same direction, i.e., the second vertex of each edge in the sequence is the first vertex of the next edge in the sequence. For example, in the directed graph illustrated above, the sequence $\langle SS, fe \rangle \langle fe, fo \rangle$ is a path from SS to fo , and $\langle SS, fo \rangle$ is another path from SS to fo . We will usually write such paths as $SS \rightarrow fe \rightarrow fo$, and $fe \rightarrow fo$,

respectively. A path will be said to **contain** a variable or vertex if that vertex occurs in one of the directed edges in the path. A path that contains a subpath beginning and ending in the same vertex is said to contain a **cycle** or to be **cyclic**. A path that does not contain a cycle is **acyclic** or **open**. A directed graph that contains a cyclic path is said to be cyclic ; a directed graph that does not contain any cyclic path is said to be acyclic.

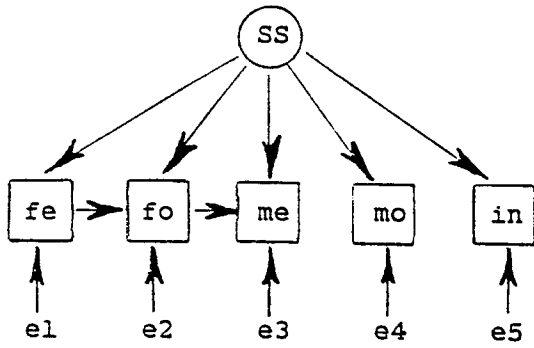
A path in a causal model represents a hypothetical causal influence that runs from one variable to another. Such an influence produces a correlation or covariance between the variables that are linked by the path. A model that postulates a path connecting variables that are actually correlated thus provides a means to explain the correlation.

A causal model can also explain a correlation between two variables as the result of other variables that affect both of them. The model depicted above predicts a correlation between *me* and *fo* because each of these variables is assumed to be affected by a third variable, *SS*. We call this sort of connection between two variables a **trek**, and we understand a path to be just one special sort of trek.

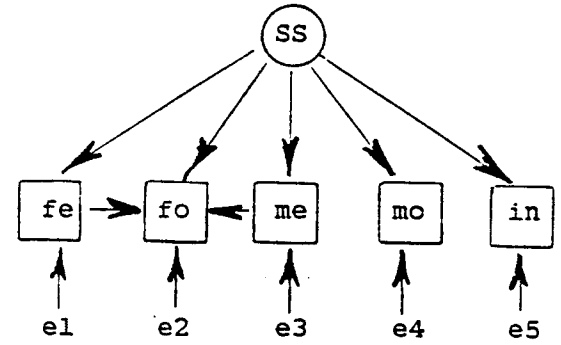
More exactly, a **trek** between two variables *u* and *v* is either a path from *v* to *u*, or a path from *u* to *v*, or a pair of paths from some variable, *w*, to *u* and to *v* such that the two paths have exactly one variable, *w*, in common. An **open trek** is a trek in which each path is acyclic. The unique variable common to both paths in an open trek is called the **source** of the trek.

For example, in the directed graph above, the pair of paths $SS \rightarrow fo$ and $SS \rightarrow mo$ is a trek between *fo* and *mo*. *SS* is its source. The pair of paths $SS \rightarrow fe \rightarrow fo$ and $SS \rightarrow mo$ is another trek between *fo* and *mo*, and *SS* is again its source. The path $fe \rightarrow fo$ is a trek with *fe* as its source.

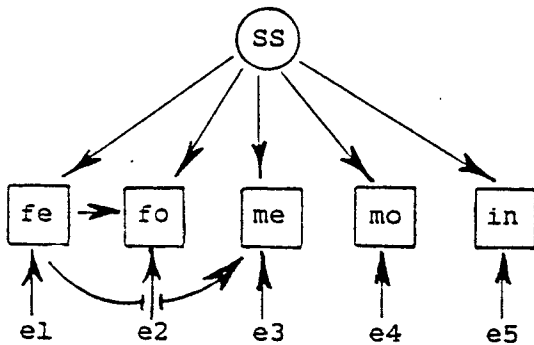
There are a variety of ways to introduce new treks between variables. Suppose we want to introduce a new trek between *fo* and *me* in Figure 4-1. We could do this by adding a directed edge $fo \rightarrow me$ (Figure 4-2-a), or a directed edge $me \rightarrow fo$ (Figure 4-2-b), or a directed edge $fe \rightarrow me$, (Figure 4-2-c). We can even add a new variable, call it *G*, and introduce the directed edges $G \rightarrow fo$ and $G \rightarrow me$, (Figure 4-2-d), and this list is by no means exhaustive.



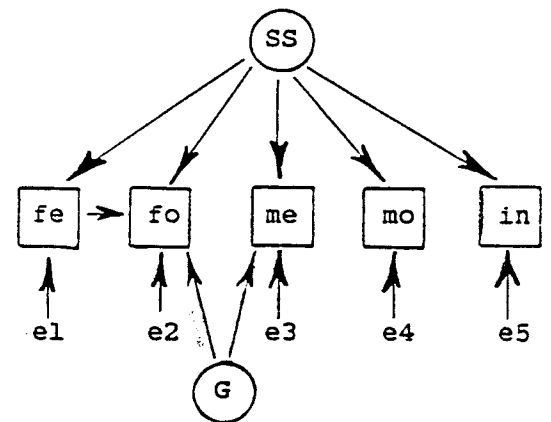
(a)



(b)



(c)



(d)

Figure 4-2: Alternative Ways to Add a Trek

Finally, we can include a correlation that is not accounted for by a causal model simply by postulating an unexplained correlation between the appropriate variables, or between causes of those variables that are external to the causal model. Such external sources of variance are often called error variables. We can represent unexplained correlations of error variables by including the error variables in the directed graph and drawing an undirected line between them, as we do in Figure 4-3.

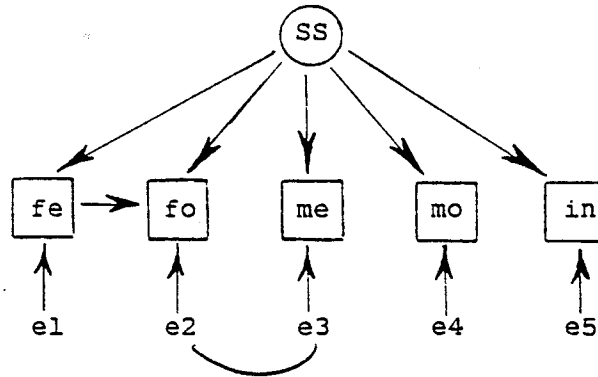


Figure 4-3: Producing a New Trek by Correlating Error Terms

The correlated errors e_2 and e_3 create a new trek between fo and me , even though such connections are not explicit in the definition of "trek" just given. The calculation of treks in models with correlated errors is described in the next section.

Different ways of adding directed edges to an initial model may have differing statistical effects because the total set of treks created may be different. In Figure 4-2-b we add $me \rightarrow fo$ to our original model and no new trek (and hence no new source of correlation) is created between fe and me . In Figure 4-2-a we add $fo \rightarrow me$ to our model and we also create a new trek (and a new source of correlation) between fe and me because of the link $fe \rightarrow fo \rightarrow me$. These are statistically inequivalent causal models. Different ways of adding directed edges and unexplained error correlations can also have differing statistical effects. In Figure 4-3 a correlated error is introduced between fo and me . This model is also statistically inequivalent to the model of Figure 4-2-a. Such statistical inequivalences prove to be extremely important in guiding the search for alternatives to an initial model.

The number of directed edges in a model is important statistically because each additional edge reduces the number of degrees of freedom of the model by one. The same is true if instead of adding new treks, we add unexplained correlations of error variables. If we add an unexplained correlation of the errors of fo and me we do not introduce a new source of correlation between fe and me . But we do reduce by one the degrees of freedom of the model.

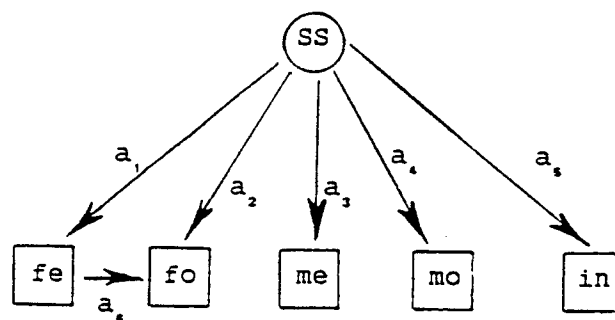
4.4. STATISTICAL MODELS FROM CAUSAL MODELS

Causal models form part, often a tacit part, of statistical models. To expand a causal model into a statistical model, equations must be written specifying the relations among the variables, and appropriate stochastic assumptions must be made. Since the work of Herbert Simon [86, 87], considerable thought has been given to how the causal part of a model can be extracted from a set of equations and stochastic assumptions. In practice, a stylistic convention is often adhered to: writers give an affected variable on the left hand side of an equation and its direct causes on the right hand side of an equation. With this convention in mind, one can often infer the directed graph a theorist intends from the equations given. This kind of reconstruction of the causal model from the equations is not, however, the sort of reduction of causal hypotheses to properties of equation systems that Simon was after, since it depends on a stylistic convention. We are concerned with the reverse problem: Starting with a causal model, which is merely a directed graph, how are the equations of a statistical model to be obtained? For that, either of two simple processes will suffice.

Method I

To get a statistical model from a directed graph simply do the following:

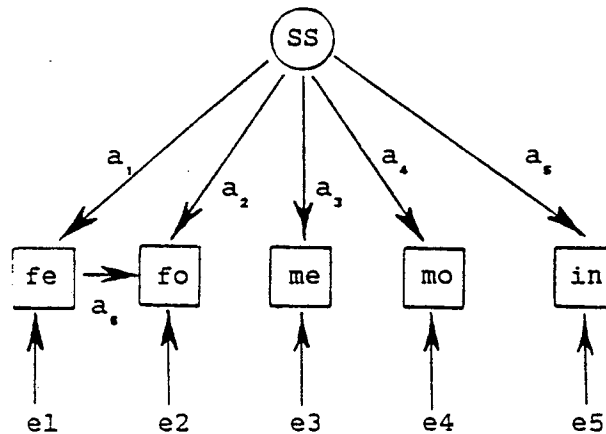
1. Attach a distinct parameter to each directed edge in the directed graph of the model. For example:



2. For each variable in the directed graph that has a causal predecessor, (i.e., the variable is not of zero indegree) add an "error" variable directed into that variable but not into any

other variable.

Thus the model above becomes:



3. For each variable in the model that is not of zero indegree, write an equation specifying that variable as a linear function of the variables directed into it. The parameters are the linear coefficients. We have then:

$$fe = a_1 SS + e_1$$

$$fo = a_2 SS + a_6 fe + e_2$$

$$me = a_3 SS + e_3$$

$$mo = a_4 SS + e_4$$

$$in = a_5 SS + e_5$$

4. Assume that any pair of variables that are *not connected by a trek* are statistically independent.

5. Specify that the linear coefficients are either constants of undetermined value, or have a determinate value, or are random variables distributed independently of each other and of all other variables in the model.

6. Consistent with the above, make whatever assumption about the distribution of the random variables seems appropriate.

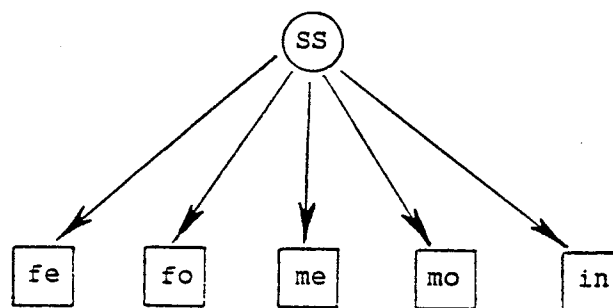
Method II

The models obtained by Method I all have uncorrelated errors. Since there are no treks between the e variables, they are all assumed to be uncorrelated with one another. That assumption is unnecessary. Instead, when a full statistical model is constructed, follow steps 1 through 3 of Method I, but rather than assuming that the errors are uncorrelated, postulate a correlation of the error variables directed into the two variables.

When correlated errors are introduced into a model, the treks between variables are calculated as if a correlated error between two variables were really a new unmeasured common cause affecting both variables.

In the same way, if a model has two latent variables, and no causal connection between them is specified, but it is assumed that the latent variables are correlated, the treks are computed as if the latent variables had a common, latent cause.

Finally, consistent with these conventions, postulate that variables not connected by a trek are statistically independent. For example, suppose one has a causal model like that illustrated but with a correlated error between fe and fo rather than the directed edge from fe to fo . Then simply take the initial model:



and proceed as in steps 1, 2 and 3 to obtain Figure 4-4.

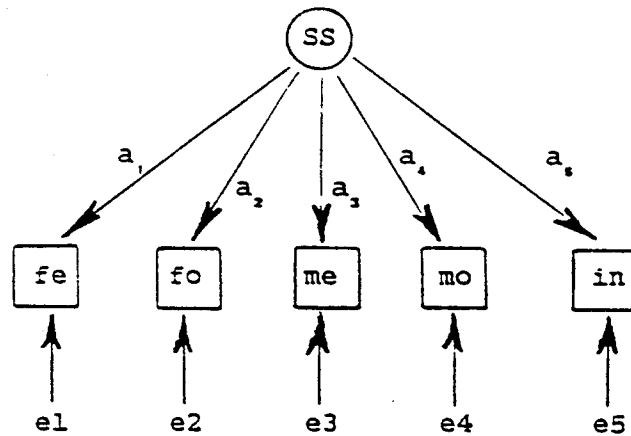


Figure 4-4:

Finally, specify that the variables e_2 and e_1 are correlated. Pictorially, the final model can be represented by drawing an *undirected* line¹² connecting e_2 and e_1 , as we do in Figure 4-5.

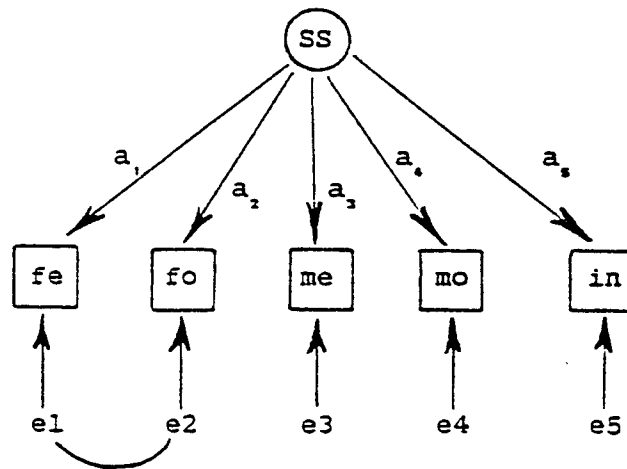


Figure 4-5: Statistical Model With Correlated Error Terms

The procedures in the TETRAD program assume that if a full statistical model is associated with a causal model suggested or analyzed by the program, then the statistical model can be obtained in one of the ways just described. In a later chapter we will describe how to turn TETRAD models into models suitable for estimation and statistical testing by standard

¹²Some writers use an "arrow" with two heads and no tail to signify such correlations

packages such as LISREL and EQS.

If the linear coefficients are assumed to be constants and not random variables, then the models that result from a directed graph through the procedures just described are **structural equation models**.¹⁴ So far as the TETRAD program is concerned, many of the assumptions usually made in structural equation modelling can be considerably relaxed. The program will work for models in which the linear coefficients are not constant, but instead are independently distributed random variables. It applies whether or not the error distributions for different members of the sample have the same probability distribution, so that models in which the variance of the error variables is not constant (heteroscedastic models) are permitted. The variables of the model can be standardized or not.

4.5. TREKS AND COORDINATING PATH EFFECTS

The **covariance** of two random variables, X and Y, is just the difference between the expected value of the product XY and the product of the expected value of X and the expected value of Y, that is $\text{Exp}(XY) - \text{Exp}(X)\text{Exp}(Y)$. The **correlation** of X and Y will be their covariance divided by the product of their standard deviations. We assume that the random variables always have distributions so that these statistics are well defined. The data for which causal models are to account are usually covariances or correlations, and it is data of this kind that the TETRAD program is meant to address.

The first job of a theory is to explain the data. In the case of linear statistical models, that usually means explaining the correlations or covariances among the measured variables in some sample. Suppose we have a statistical model, and the linear coefficients in that model are somehow specified, whether by statistical estimation, or a priori, or however.¹⁵ The estimated statistical model does not determine unique values for the covariances of the measured variables in the population it is to describe, nor does it determine probability distributions for covariances in a sample drawn from the population.

To calculate population covariances we need to know the variance of every variable of zero indegree, including the error variables, and we need to know the covariance of any

¹⁴ The stochastic assumptions that usually accompany structural equation models can be understood in the following way: The structural equations are understood to apply to each member of the sample. Thus the equations above should be indexed by i , where i denotes the i th member of the sample. The linear coefficients are not random variables and are assumed (usually contrary to fact) to be constant in the population. e_{1i} is assumed to be statistically independent of e_{1j} if i is not equal to j . Usually, e_{1i} and e_{1j} are assumed to have the same probability distribution. Similarly for e_{2i} , e_{2j} , etc. These assumptions characteristic of structural equation modelling are appropriate if the stochastic elements are attributed to random sampling with replacement.

¹⁵ In the case of models with stochastic coefficients, assume the population mean value is specified.

pair of variables that have, according to the model, an unexplained correlation. *These variances and covariances, together with the directed graph and the values of the linear coefficients, determine the population covariances of the measured variables according to the model.*

Return to the simple model we considered earlier.

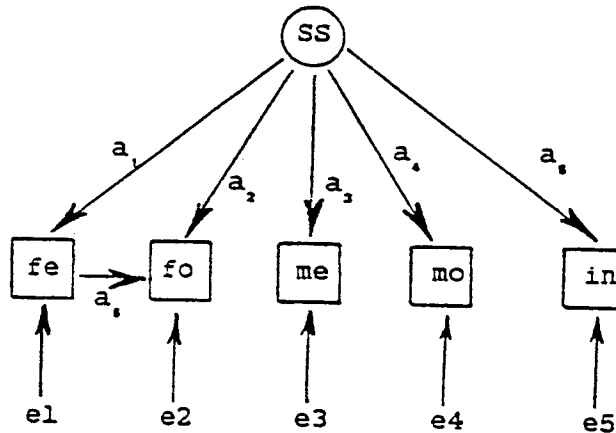


Figure 4-6:

One trek between fe and fo is $\langle SS, fe \rangle, \langle SS, fo \rangle$ with coefficients a_1 and a_2 , and the other is $\langle fe, fo \rangle$ with coefficient a_6 . With any path in the directed graph of a model, we associate the product of the linear coefficients associated with the graph. Thus the path $\langle SS, fe \rangle, \langle fe, fo \rangle$ is associated with the product $a_1 a_6$. The coefficient product is unique up to permutation of the coefficients in the product; no two distinct paths have the same product of coefficients. Since a trek consists of at most two paths, we can also associate each trek with a product of linear coefficients, namely the product of all of the coefficients for any edge in each path in the trek. Thus the trek $\langle SS, fe \rangle, \langle SS, fo \rangle$ has the product $a_1 a_2$. Again, the product of coefficients associated with any trek is unique up to permutation of the factors, and no two distinct treks have the same product of coefficients.

A coordinating path effect for two variables X, Y having a trek with source S connecting X and Y is the product of the trek coefficients multiplied by the variance of S . For example, the coordinating path effect for the trek $\langle SS, fe \rangle, \langle SS, fo \rangle$ in the preceding example is just $a_1 a_2 \text{Var}(SS)$.

For any graph, the covariance of two variables, X, Y is just the sum, over all distinct treks between X and Y, of the coordinating path effects.

The covariance of f_e and f_o , $\text{Cov}(f_e, f_o)$, in the preceding example is just $a_1 a_2 \text{Var}(SS) + a_6 \text{Var}(f_e)$. In applying the rule, source variables that have unexplained covariances postulated by the model should simply be viewed as having a common latent cause¹⁶ whose variance is equal to the unexplained covariance.

So the population covariances a model implies are determined by the variances of its source variables, the linear coefficients, and the directed graph of causal relations postulated by the model. The probability distribution of the sample covariances is determined by these factors plus the joint distribution assumed for the variables. The directed graph itself determines very little about the covariances. In fact, all that the directed graph may imply about the values of any particular covariance is that, if the covariates are not connected by a trek, then their covariance is zero. But if we are to search for alternative models to explain the data, and the automatic part of the search uses only the data, the directed graphs of alternative models, and some distribution assumptions, then the directed graph *alone* must determine something about the population covariances. And it does. *The directed graph of a model determines patterns among the covariances of the measured variables in the population, without determining unique values for those covariances.* These patterns are overidentifying constraints.

4.6. CONSTRAINTS ON CORRELATIONS

4.6.1. Overidentifying Constraints and Directed Graphs

Before the coefficients of a statistical model are estimated or specified, the model does not imply unique values for the population covariances, but it can imply that whatever those covariances may be, they must satisfy certain conditions. Whatever the population covariances may be, if they are consistent with the model then they will satisfy these conditions, no matter what values the linear coefficients may have. Conditions of this kind are called **overidentifying constraints** in the econometrics literature, for the reason that when a model implies such constraints there are generally multiple ways to estimate its parameters. Blalock [7] calls these conditions **prediction equations**, because they also represent implications of a statistical model that can be tested more or less directly.

¹⁶That is, if X and Y are assumed to have correlated errors, in applying this rule one treats them as though the model contained a latent variable T and directed edges $T \rightarrow X$ and $T \rightarrow Y$ whose associated coefficients are fixed at one.

Certain important kinds of overidentifying constraints are determined entirely by the directed graph of a model. In order to determine whether a statistical model implies these sorts of constraints, one need only examine the causal model embedded in it. Although there are other kinds, the two sorts of overidentifying constraints that we shall consider, and that are entirely determined by the directed graph, are **vanishing partial correlations** and **tetrad equations**.

4.6.2. Vanishing Partial Correlations

Consider the simple example given by the following causal model and associated structural equation model:

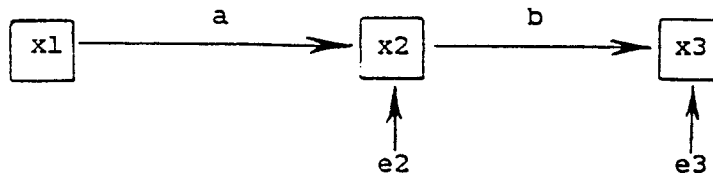


Figure 4-7:

Given any arbitrary numbering of a set of measured variables, the correlations of the variables can be arranged into a square, symmetric matrix whose diagonal elements are all unity. If the correlations are those that obtain in the population, the matrix is called the **population correlation matrix**. If the correlations are those in a sample, then the matrix is the **sample correlation matrix**. In the present example we can write the population correlation matrix as:

$$P = \begin{pmatrix} \rho_{11} & \rho_{12} & \rho_{13} \\ \rho_{21} & \rho_{22} & \rho_{23} \\ \rho_{31} & \rho_{32} & \rho_{33} \end{pmatrix} = \begin{pmatrix} 1 & a & ab \\ a & 1 & b \\ ab & b & 1 \end{pmatrix}$$

where a and b are undetermined linear coefficients.

In this simple example the model implies a constraint on the correlation matrix. The constraint is easily obtained by considering the treks in the model. First, however, we will derive the constraint in the conventional way, by means of the partial correlation

coefficient. The partial correlation of x_1 and x_3 with respect to x_2 is defined to be (see Anderson, [2]):

$$\rho_{13.2} = \frac{\rho_{13} - \rho_{12}\rho_{23}}{(1 - \rho_{12}^2)^{1/2} (1 - \rho_{23}^2)^{1/2}}$$

which in this case

$$= \frac{ab - ab}{(1 - a^2)^{1/2} (1 - b^2)^{1/2}} = 0$$

Thus the model implies that

$$\rho_{13} - \rho_{12}\rho_{23} = 0$$

which is a constraint on the correlations.

Note that the derivation of this constraint did not depend on assigning any particular numerical values to the coefficients a and b . The model will imply the vanishing partial correlation *no matter what* values a and b might have.¹⁷ We are especially interested in models that imply constraints no matter what values the non-zero linear coefficients may have. We will say in such cases that the model **implies** the constraint, or **entails** the constraint, or sometimes for emphasis that the model **strongly implies** or **strongly entails** the constraint. In cases in which estimated models determine a constraint because of particular values of the parameters, we will say that such a model **weakly implies** the constraint.

We can derive the same constraint by considering the treks in the graph of the model: If the variables are standardized, the correlation between two measured variables is equal to the sum, over all treks connecting the variables, of the product of the coefficients corresponding to the directed edges in the trek (see Heise [43]). Consider the graph of the model again:

¹⁷Unless, of course, $a = b = 1$, in which case the partial correlation is not defined.

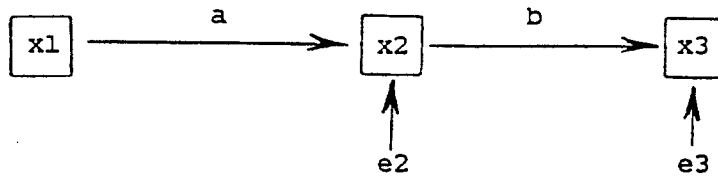


Figure 4-8:

In this case there is but one trek between x_1 and x_3 , hence:

$$\rho_{13} = ab = \rho_{12}\rho_{23}$$

or

$$\rho_{13} - \rho_{12}\rho_{23} = 0.$$

It turns out that this derivation does not depend on the assumption that the variables are standardized. We do not need to know or stipulate anything about the variances of the variables in order to derive the vanishing partial correlation from the directed graph. There is a general principle involved in the example.

Theorem 1: The *directed graph* of a model *alone* determines the vanishing partial correlations the model implies.

This theorem (as well as the remaining theorems stated in this chapter) holds no matter what the graph may be. It holds for cyclic graphs as well as acyclic graphs. In Chapter 10 we prove the following theorem, which provides a condition on directed graphs that is necessary and sufficient for the implication of a vanishing partial correlation:

Theorem 2: For any directed graph, G , and any three distinct variables, x , y and z which are vertices of G , the following two conditions are equivalent

- (1) every trek between x and z contains y , and either every trek between y and z contains an edge directed out of y or every trek between x and y contains an edge directed out of y ;

(2) any model for which G is the causal graph implies that $\rho_{xz} - \rho_{xy}\rho_{yz} = 0$.

This theorem means that the partial correlation constraints implied by a linear causal model can be determined entirely from its directed graph. For example the directed graph:



(where we neglect to write the error variables explicitly) entails that

$$\rho_{24} - \rho_{23}\rho_{34} = 0$$

$$\rho_{23} - \rho_{23}\rho_{33} = 0$$

as well as other vanishing partial correlations. The partial correlation $\rho_{23.3}$ is undefined because its denominator is identically zero, but the numerator meets the conditions of the theorem if in the theorem we take x to be x_2 , y to be x_3 and z to be x_3 and note that there is in that case no trek between y and z , so all treks between y and z (of which there are none) satisfy the condition that they contain an edge out of y .

There are fast algorithms for determining the properties of directed graphs that are relevant to the implication of vanishing partial correlations, and the TETRAD program embodies such an algorithm. In Chapter 10 we give the algorithm and prove its adequacy.

4.6.3. Tetrad Equations

Tetrad equations involve products of correlations among a set of four measured variables. There are three possible inequivalent products of two correlations which involve all and only four variables. Eg., if the four variables are w , x , y , and z , then the three possible correlation products are

$$\begin{aligned} &\rho_{wx}\rho_{yz} \\ &\rho_{wy}\rho_{xz} \\ &\rho_{wz}\rho_{xy} \end{aligned}$$

Tetrad equations say that one product of correlations (or covariances) equals another product of correlations (or covariances). There are three possible tetrad equations, any two of which are independent, in a set of four variables. They are:

$$\begin{aligned}\rho_{wx}\rho_{yz} &= \rho_{wy}\rho_{xz} \\ \rho_{wy}\rho_{xz} &= \rho_{wz}\rho_{xy} \\ \rho_{wx}\rho_{yz} &= \rho_{wz}\rho_{xy}\end{aligned}$$

The simple model:

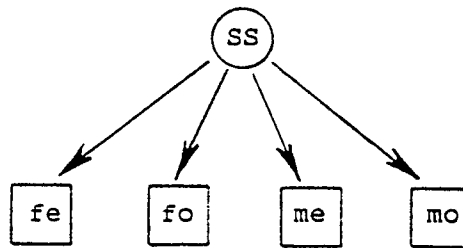


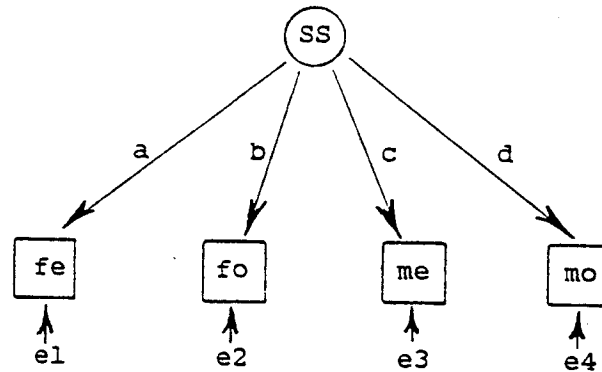
Figure 4-9: One Factor model

entails the tetrad equations:

$$\begin{aligned}\rho_{fo,fe}\rho_{me,mo} &= \rho_{fo,me}\rho_{fe,mo} \\ \rho_{fo,mo}\rho_{fe,me} &= \rho_{fo,fe}\rho_{me,mo} \\ \rho_{fo,mo}\rho_{fe,me} &= \rho_{fo,me}\rho_{fe,mo}\end{aligned}$$

More complex models can entail much more complicated sets of tetrad equations.

In simplest terms, a tetrad equation is implied by a causal model if and only if the directed graph associated with the model possesses an appropriate symmetry property. Suppose we give each directed edge in Figure 4-9 a distinct letter as a label. For example:



Suppose we traverse the treks between measured variables, forming the product of the labels of all of the directed edges we pass over. In this case there is but one trek between each pair of variables. If we follow the trek connecting *fe* and *fo* we pass across the directed edge from *SS* to *fe*, with label *a*, and across the directed edge from *SS* to *fo*, with label *b*. So we form the product of labels, (*ab*), and for reasons that are explained in detail in Chapter 10, we associate that product of labels with the correlation, $\rho_{fe,fo}$, of *fe* and *fo*. In the same way, we associate the correlation of *me* and *mo* with the the product (*cd*), of *fe* and *me* with the product (*ac*), of *fe* and *mo* with the product (*ad*), and so on. This labeled graph has the following symmetry: If we multiply the label product (*ab*) by the label product (*cd*), we get the same total product that is obtained by multiplying the label product (*ad*) by the label product (*bc*):

$$(ab)(cd) = (ad)(bc)$$

On substituting the associated correlations for the respective label products, this equation becomes a tetrad equation:

$$\rho_{fe,fo}\rho_{me,mo} = \rho_{fe,mo}\rho_{fe,me}$$

In the same way, the two other tetrad equations implied by this model can be derived from examining the symmetry of label products.

Theorem 3: The tetrad equations strongly implied by a statistical model are determined entirely by the directed graph of the model.

The tetrad equations implied by a causal model are therefore also implied by all statistical models whose causal hypotheses are represented by the causal model. The causal model makes certain assumptions about the joint distribution of some variables, and about the existence of moments, but does not otherwise restrict the distribution of each variable individually. How the variables are distributed does not matter: the tetrad equations are implied no matter whether the variables are standardized or unstandardized, normally distributed or not.

In fact something much stronger than Theorem 3 holds. There exists an algorithm for deriving the tetrad equations implied by *any* directed graph. The algorithm is described in chapter 10, where its adequacy is proved.

4.6.4. Tetrad Equations Derived from Vanishing Partial Correlations

We have said that the implication of both vanishing partial correlations and of tetrad equations is determined by the directed graph of a model, and we have said, in Theorem 2, what property of a directed graph determines whether or not that graph implies a particular vanishing partial correlation. We have noted the symmetry property that determines tetrad equations, and that property in fact gives an entirely graph theoretic means of computing the tetrad equations implied by a model. One might wonder whether there is some connection between vanishing partial correlations and tetrad equations. There is. A directed graph implies a tetrad equation if it implies that four numerators of partial correlations vanish:

Theorem 4: A directed graph G , implies the tetrad equation

$$\rho_{ij}\rho_{kl} = \rho_{ik}\rho_{jl}$$

if for some variable v it implies the four conditions:

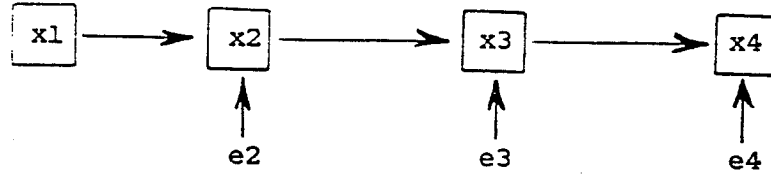
$$\rho_{ij} - \rho_{iv}\rho_{jv} = 0$$

$$\rho_{kl} - \rho_{kv}\rho_{lv} = 0$$

$$\rho_{ik} - \rho_{iv}\rho_{kv} = 0$$

$$\rho_{jl} - \rho_{jv}\rho_{vl} = 0.$$

For example, the model :



implies that

$$\rho_{13}\rho_{24} = \rho_{14}\rho_{23}$$

and the reader can verify that the numerators of the partial correlations

$$\rho_{13.3}$$

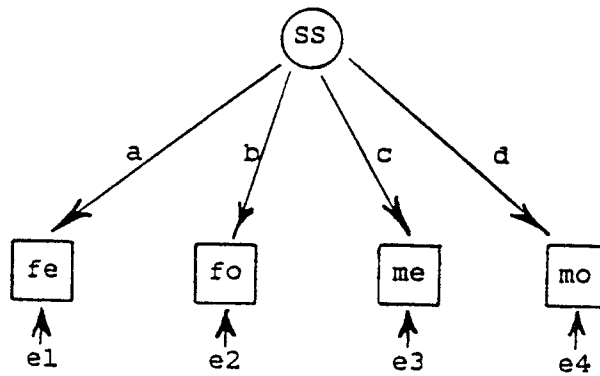
$$\rho_{24.3}$$

$$\rho_{14.3}$$

$$\rho_{23.3}$$

must all vanish.

In the model



all three possible tetrad equations among fe, fo, me and mo are implied, because the model

implies that all of the partial correlations among these variables vanish when SS is controlled for.

The converse of this theorem is not true. A model can imply a vanishing tetrad difference without implying vanishing partial correlations that in turn imply the tetrad equation. For example, the model (in which we neglect to exhibit error terms):



implies that $\rho_{15}\rho_{23} - \rho_{25}\rho_{13} = 0$, but it does not imply any corresponding set of vanishing partial correlations.

4.6.5. Graphs and Constraints

Because of the preceding theorems it makes sense to speak of the vanishing tetrad differences and vanishing partial correlations as implied by a graph. A few elementary definitions enable us to state several important properties of graphs and the constraints they imply. A graph can be described by its set V of vertices and its set E of edges. $G = \langle V, E \rangle$ is completely disconnected if and only if it has no edges, i.e., if and only if E is empty. G' is a subgraph of G if and only if $V' \subseteq V$ and $E' \subseteq E$. With these definitions, the following facts hold.

Theorem 5: A completely disconnected graph implies all possible vanishing tetrad differences and vanishing partial correlations among its vertices.

Theorem 6: If G' is a subgraph of G , then G' implies every vanishing tetrad difference and every vanishing partial correlation implied by G .

4.6.6. Other Constraints

There are many other forms of constraints that can be implied by linear models. Models with six or more measured variables can imply vanishing sextad differences, for example, and models with eight or more measured variables can imply vanishing octad differences, and so on. A model only implies higher order constraints of this kind, however, if it also

implies a set of vanishing tetrad differences that also imply the higher order vanishing differences. So in a sense, these higher order constraints are nothing new.

A model may also imply higher order vanishing partial correlations, in which the effects of two or more variables are controlled for. TETRAD does *not* compute higher order vanishing partial correlations that may be implied by the model.

There are in addition higher order constraints that are independent of tetrad constraints. A model with at least five measured variables may imply **pentad** constraints, which assert that a sum of products vanishes, where each product has five correlations as factors. If a model implying pentad constraints has but one common source, then the pentad constraints will be implied as well by the tetrad equations the model implies, but other models can imply pentad constraints that are independent of vanishing tetrad differences.

The existence of other constraints means that a procedure, such as TETRAD, that searches for models by attempting to explain the vanishing tetrad differences and vanishing partial correlations found in the data, will tend to be too generous. It will generate models that do not explain higher order constraints that hold in the data, and it will also generate models that imply higher order constraints that do not hold in the data. We will see in Chapter 8 that TETRAD occasionally generates alternative models that perform very differently when tested statistically.

4.6.7. Latent Variables and Overidentifying Constraints

Nearly all causal models contain "latent" variables. Latent variables are simply any variables that occur in the model but that have not been measured. The most ubiquitous sort of latent variables are error terms, but latent variables also often occur as common causes of several measured variables, and unlike error variables they usually carry with them some theoretical interpretation.

Notationally, TETRAD distinguishes between latent and measured variables by whether or not the first character in the variable's name is capitalized. If the first letter is capitalized, TETRAD interprets it as a latent variable. In our figures, we follow the same convention. We also follow a somewhat standard convention of putting measured variables in boxes, interpreted latent variables in circles, and uninterpreted error terms in nothing at all.

"Latent" does not mean "unobservable," and merely because a variable happens not to be among those measured in a study does not mean that the the variable could not be measured, were we careful and clever enough. Some philosophical and methodological

issues concerning latent variables will be discussed in a later chapter. For now, we wish only to explain some connections between latent variables and overidentifying constraints.

We know from Theorem 4 that a model implies a tetrad equation if it implies that all of the numerators of their partial correlations vanish when taken with respect to *some* common variable. If the converse of Theorem 4 were true, then whenever we found a set of tetrad equations to hold, and also found that among the measured variables, no set of vanishing partial correlations hold that imply the tetrad equation, then we could conclude that the tetrad equations could only be explained by a model that introduces a latent variable. We know from the model shown in Figure 4-9 that the converse of Theorem 4 is *not* true. We have, however, found no case in which a model *without cycles* implies a tetrad equation unless it also implies a set of vanishing partial correlations that imply that tetrad equation. Our experience is not proof that such models do not exist, but we suggest the following two heuristics:

1) *If a set of tetrad equations hold among measured variables, and no vanishing partial correlations hold among the measured variables that imply the tetrad equations, and no cyclic model without latent variables is acceptable, then ceteris paribus a model with latent variables provides the best explanation of the data.*

2) *If a set of vanishing partial correlations among a set of measured variables hold in the data and imply tetrad equations that also hold, then do not introduce a latent variable as a common cause of the measured variables unless there is good substantive reason to do so.*

4.7. CORRELATED ERRORS ARE NOT EQUIVALENT TO DIRECT EFFECTS

Part of the power of the TETRAD program lies in its ability to use overidentifying constraints to discriminate among circumstances in which X causes Y, Y causes X or X and Y have correlated errors. Models with direct causal effects between their measured variables may have quite different statistical properties from models in which the direct effects are replaced by correlated errors. That is because the statistical implications of a causal model are determined by its set of treks, and directed edges can create very different treks than those created by correlated errors. Since this aspect of causal models, although simple enough, is sometimes misunderstood, we will give a more formal demonstration.

Two examples of linear causal models, including both graphs and equations, are given

below. Structurally, they are typical of many measurement models in sociology, psychometrics and social psychology.

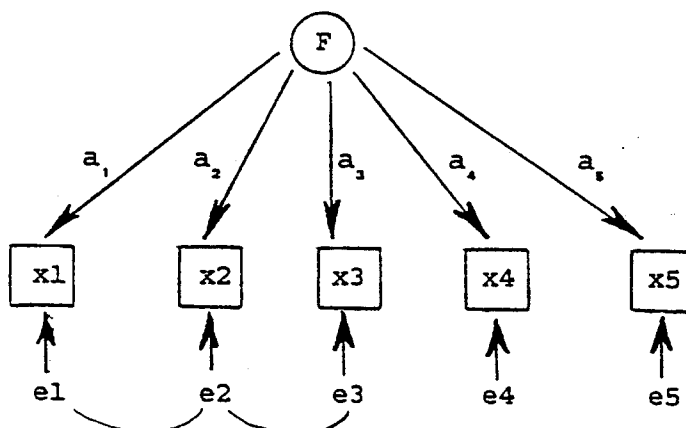


Figure 4-10: Model 1

The equations for Model 1 are:

$$x_1 = a_1 F + e_1$$

$$x_2 = a_2 F + e_2$$

$$x_3 = a_3 F + e_3$$

$$x_4 = a_4 F + e_4$$

$$x_5 = a_5 F + e_5$$

or in matrix form

$$X = A(F) + E$$

Here X is the column vector of the x_i , A is the column vector of the a_i , (F) is the matrix whose one element is F , and E is the column vector of e_i . To get the correlation matrix for this model we multiply each side of the matrix equation by its transpose and take the expectation value of both sides of the resulting matrix equation. (We assume that the means of all variables are zero and that F has unit variance).

The result is

$$\Sigma = \text{Exp}(XX^t) = A\Pi A^t + T + \text{Exp}(A(F)E^t) + \text{Exp}(E(F)A^t)$$

The graph implies (under the assumption that variables between which there is no trek are statistically independent) that the last two terms on the right hand side are zero. Intuitively, this is because the model implies that the error terms are independent of the latent factor F. Here Π is just the one element matrix consisting of the variance of F. T , the covariance matrix of the e_i variables, is a symmetric matrix with non-zero diagonal and with its (1,2), (2,1), (2,3) and (3,2) members also nonzero. The independent parameters of the model consist of the five members of A (the a_i), the variance of F, the five diagonal members of T and any two independent non-diagonal non-zero members of T . A statistical model is not yet fully determined until one specifies what family of probability distributions the parameters parameterize. If, as is commonly assumed, the distribution is multinormal, then Σ parameterizes the distributions since the means are all assumed to be zero.

The components of the preceding matrix equation assert that

$$\rho_{ij} = a_i a_j$$

Elementary algebra then shows that this model entails that

$$\rho_{13}\rho_{45} = \rho_{14}\rho_{35} = \rho_{15}\rho_{34}$$

We recognize these equations as three tetrad equations.

Consider a second model, shown in Fig. 4-11.

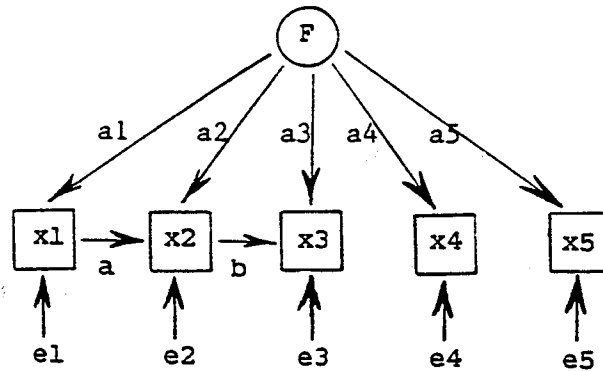


Figure 4-11: Model 2

The equations for this model are:

$$x_1 = a_1 F + e_1$$

$$x_2 - a x_1 = a_2 F + e_2$$

$$x_3 - b x_2 = a_3 F + e_3$$

$$x_4 = a_4 F + e_4$$

$$x_5 = a_5 F + e_5$$

In matrix form these equations are

$$BX = A(F) + E$$

or equivalently,

$$X = B^{-1}A(F) + B^{-1}E$$

where X is a column vector of the x_i variables, A is the column matrix of the a_i , E is the column vector of the variables e_i , and

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -a & 1 & 0 & 0 & 0 \\ 0 & -b & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The matrix B^{-1} looks like B except that $-a$ is replaced by a and $-b$ is replaced by b and the (3,1) element is not zero but ab .

The covariance matrix is again obtained by multiplying both sides of the matrix equation by their respective transposes, and taking expectation values. This gives:

$$\Sigma = \text{Exp}(XX^t) = B^{-1}A\Pi A^t B^{-1t} + B^{-1}TB^{-1t} + \text{Exp}(B^{-1}E(F)A^t B^{-1t}) + \text{Exp}(B^{-1}A(F)E^t B^{-1t})$$

where Π is the variance matrix of F , and T is the covariance matrix of the e_i variables. The last two terms on the right vanish. That is again because there is no trek between any e_i variable and F . Therefore F and each e_i must be statistically independent, which implies, since all variables have zero means, that the expected value of $F \cdot e_i$ is zero for all i . For the same reason the expected value of the product of any two distinct e_i variables is zero. All of the off-diagonal elements of T are therefore zero.

Examining the components of the covariance matrix, we find that the equation

$$\rho_{13}\rho_{45} = \rho_{14}\rho_{35}$$

is not implied, which proves that the two models imply different constraints on the covariance matrix, and are therefore inequivalent.

4.8. STATISTICAL ISSUES, BRIEFLY CONSIDERED

The parameters of a model must be estimated from sample data, and sample data must be used to perform statistical tests of a model. The TETRAD program does not do statistical estimation of parameters, and performs only statistical tests for vanishing tetrad

differences and vanishing partial correlations, but no tests of models or of the significance of parameter estimates. TETRAD is a program for helping in model specification, not for estimating parameters and performing statistical tests of models. Thus the program is best used in conjunction with a statistical package that will perform these analyses. There are several packages of this kind, some of them virtually indistinguishable save for ease of use or speed of processing.

Different statistical procedures are favored in different social science disciplines, often for reasons that are unclear. Econometricians tend to prefer least squares estimates of parameters, perhaps because they so often consider regression models. (From a graph theoretic point of view, regression models are simply those in which one dependent variable is affected by several independent variables, and there are no further causal connections or correlated errors, so that the system can be represented by a single equation.) For more complex models, multi-stage least squares procedures are often used in econometrics. In other social sciences, full information maximum likelihood estimators are more commonly used. These procedures have been automated in the LISREL computer program and in the recently released EQS computer program. In later chapters we will estimate parameters using the EQS program.

Hypothesis tests are also performed in different ways in different social science disciplines. There are, for example, various tests of the assumption of linearity but they are rarely performed in many social science disciplines in which linear models are used. A specific model can only be tested provided the model implies some constraint on the variance/covariance matrix. Regression models do not, and so the tests performed on them are usually only tests of the estimated regression coefficients. In social sciences outside of econometrics, models are often subjected to "chi square" tests in conjunction with maximum likelihood estimates of parameters. An especially clear discussion of these tests can be found in Fox, [26]. These tests compute a statistic that, on the assumption that the variable distribution is multinormal, is asymptotically distributed as chi square with the number of degrees of freedom equal to the number of variances and covariances among the measured variables minus the number of free parameters (including variances) of the model. Chi square statistics of this kind are computed by both the LISREL and EQS programs.

The proper use and value of these statistical tests is a controversial matter which we will discuss in Chapter 5.

1000
 900
 800
 700
 600
 500
 400
 300
 200
 100
 0

5. THE STRUCTURE AND METHOD OF TETRAD

5.1. INTRODUCTION

The TETRAD program helps to search for good models of correlation or covariance data. This chapter describes the methodological principles behind TETRAD, how they are realized in the program, and how these principles can be used with the program in searching for good causal models. It lists the output the program provides, describes some of the calculations that go into the output, and illustrates of how the output can be used to find good causal models. It also discusses goodness of fit tests and compares TETRAD's search strategy with other search strategies. More detailed mathematical characterizations of the fundamental algorithms of the program, and proofs of their adequacy, are given in chapter 10 of this book. Several cases illustrating the use of the program on empirical and simulated data are described in later chapters.

5.2. THE METHODOLOGICAL PRINCIPLES THAT UNDERLIE TETRAD

5.2.1. The Inference Problem

Suppose we have a set of covariances or correlations, and an initial linear causal model to account for them. The initial model may contain latent variables, but it is otherwise very simple in the sense that it postulates very few direct causal connections in order to account for the measured correlations. More technically, let us suppose that the initial model contains one and only one trek for each measured non-zero covariance. If the model is estimated from the correlations or covariances and tested, say by a chi square test, it may do very poorly, and may in fact be rejected by conventional standards. Even so, suppose we have reason to think that the model is plausible, and that it may well be true but incomplete. There may be other causal connections among the variables, there may be correlated errors, or there may be omitted unmeasured variables that have an effect on several of the measured or unmeasured variables. If the initial model yields a chi square value with zero probability, but we think its positive causal claims are nonetheless plausible, then the problem is to find the "best" elaborations of the initial model.

But what does "best" mean? No one has a complete answer to that question, but we

propose a mixture of considerations about scientific explanation:

Other things being equal, the best models are those that imply patterns or constraints judged to hold in the population, that do not imply patterns judged not to hold in the population, and that are as simple as possible.

For brevity, we will sometimes say that constraints that are judged, on the basis of a sample, to hold in the population are constraints that **hold in the data**.

5.2.2. Implying Constraints That Hold in the Data

By **Spearman's Principle** we mean the following:

Other things being equal, prefer those models that, for all values of their free parameters (the linear coefficients), entail the constraints judged to hold in the population.

This is a principle stated *ceteris paribus*: it can be defeated if one has reason to think that a constraint on the correlation matrix is satisfied by chance, or if some model that violates the principle has other extraordinary virtues, or if one has reason to think that the general assumptions of linear modeling are inappropriate to the case. The reason we call this Spearman's principle are given in chapter 9.

5.2.3. Not Implying Constraints That Do Not Hold in the Data

By **Thurstone's Principle**, we mean the following:

Other things being equal, a model should not imply constraints that are not supported by the sample data.

We call this idea Thurstone's Principle, although we mean no historical attribution, because it is derived from a principle that is fundamental to factor analysis. In factor analysis, one reduces the difference between predicted and observed correlations (the residuals) by introducing additional causal structure. The heuristic we use to reduce the residuals is to avoid implying constraints that are not supported by the data.

5.2.4. The Simplicity Principle

Nearly every methodologist, whether philosopher, statistician or scientist, has emphasized the importance of the virtue of simplicity. It is surprising that something so widely admired should be so difficult to characterize, but there are no satisfactory general accounts of the notion of simplicity that apply to the variety of scientific contexts in which the notion is used. In the special case of causal models, we understand simpler models to be those that posit fewer causal connections. In effect, we suppose an initial bias against causal connections, and require that a case be made for any causal claims. This requirement has nothing to do with whether the data are experimental or non-experimental; in fact, it is because of this bias that good evidence and arguments are demanded for controversial causal claims, whether about the efficacy of a drug or the effects of pornography.

5.2.5. The Principles Conflict

In practice, the three principles interact and conflict. Spearman's principle will tend to be satisfied by simple models, since such models imply a great many constraints. But constraints are seldom *exactly* satisfied in any sample, and thus on the average the fewer constraints implied, the better Thurstone's principle will be realized. Modifying a model to account for more and more of the empirical covariances generally means adding more and more free parameters, and thus increasing the complexity of the model, and losing degrees of freedom. The TETRAD program provides information about how modifications to an initial model affect both Spearman's and Thurstone's principles. In effect, the user is free to judge how the explanatory principles are to be weighted.

5.3. HOW THE METHODOLOGICAL PRINCIPLES ARE REALIZED IN TETRAD

5.3.1. The Constraints That TETRAD Calculates

To satisfy Spearman's or Thurstone's Principle strictly, we would have to search for the models that imply all of the constraints that are judged to hold in virtue of the sample correlations. In practice, we cannot do that. First, we do not know how to test for all forms of constraints, and second we do not have fast algorithms for computing all forms of constraints implied by an arbitrary model. Evidently, what is required is some *heuristic* procedure for searching and testing the vast space of elaborations of the initial model.

Our strategy is to look only at constraints of two special kinds: vanishing tetrad differences and vanishing partial correlations. These constraints are numerous enough that they help to discriminate between alternative elaborations of an initial model, they are rapidly computable implications of models, and statistical tests are available for them. Further, these two kinds of constraints have nice *inheritance* properties that can be used in building up the best models from the initial model. Adding directed edges or correlated errors never increases the number of vanishing tetrad differences and vanishing partials that are implied by a model. Thus if at any stage a model does not adequately meet Spearman's criterion, no further elaboration of that model will either, and we can avoid considering all such elaborations.

5.3.2. Comparing Models

TETRAD will, upon request, print a table listing every tetrad or vanishing partial equation that is implied by a model and every equation that holds to within a user set significance level. Each time a model implies an equation judged to hold in the population, the model satisfies Spearman's principle for that equation. TETRAD automatically calculates the number of equations that hold at a particular significance level, the number of those equations that are implied by the model (the number explained), and the number of equations that are judged to hold but that are *not* implied. According to Spearman's Principle, other things being equal we prefer models that explain a greater number of equations that are judged to hold in the population.

If a model implies an equation that is judged not to hold in the population, then the model *violates* Thurstone's Principle for that equation. TETRAD automatically calculates the number of equations a model implies that do not hold to within a user set significance level. Some equations violate Thurstone's principle more severely than others, i.e they have larger residuals than others, so it is not only the *number* of equations that violate Thurstone's Principle that matters, but the *total residual*. TETRAD automatically sums the residuals of all the equations implied by a model and reports that number as the **Total Tetrad Residual (TTR)** for tetrad constraints and as the **Total Partial Residual (TPR)** for vanishing partial correlation constraints. According to Thurstone's Principle, other things being equal we prefer models that have the *lowest* TTR and the *lowest* TPR.

5.3.3. Comparing Elaborations of a Model

The information TETRAD gives is based on the inference problem we sketched above: starting with a simple model, what are its best elaborations.

In order to help the user compare elaborations of a model, TETRAD supplies information on the effect each elaboration has on the model's ability to satisfy Spearman's Principle and Thurstone's Principle. If an initial model implies an equation that holds in the data, but the addition of an edge (or correlated error) would result in a model that no longer implies that equation, adding the edge causes the model to violate Spearman's Principle for that equation. If a model implies an equation that would no longer be implied when an edge is added to the model, we say that the edge **defeats** that equation. TETRAD supplies a measure of the amount of damage to Spearman's Principle each elaboration causes: **I(H-I)**. Mnemonically, this expression stands for the Increase in the number of equations that Hold but that are not Implied. Other things being equal, we prefer the elaboration that has the *lowest* I(H-I) value.

Symmetrically, if an elaboration defeats an equation that is judged not to hold in the data, then that elaboration prevents the model from violating Thurstone's Principle for that equation. The model previously implied an equation considered false, but after the elaboration it does not imply that equation. TETRAD supplies two measures of how much each elaboration helps a model to satisfy Thurstone's Principle. It calculates **D(I-H)**, the Decrease in the number of equations that are Implied that do not Hold, and it calculates an elaboration's **Rttr**, or the Reduction in the total tetrad residual that would result from the elaboration. Other things being equal, we prefer the elaboration that has the *highest* Rttr value and the *highest* D(I-H) value.

5.4. A SEARCH STRATEGY FOR FINDING GOOD CAUSAL MODELS

The three principles suggest a strategy for model specification: beginning with a very simple model, determine how it fares by Spearman's and Thurstone's criteria (number of equations explained and the TTR, respectively). Then, step by step, examine increasingly complex elaborations of the initial model, and determine how they behave. Stop elaborating as soon as it is judged that a model satisfies Spearman's and Thurstone's principles sufficiently well. In this way, unnecessarily complex models will not be considered. The steps in the search look like a tree diagram:

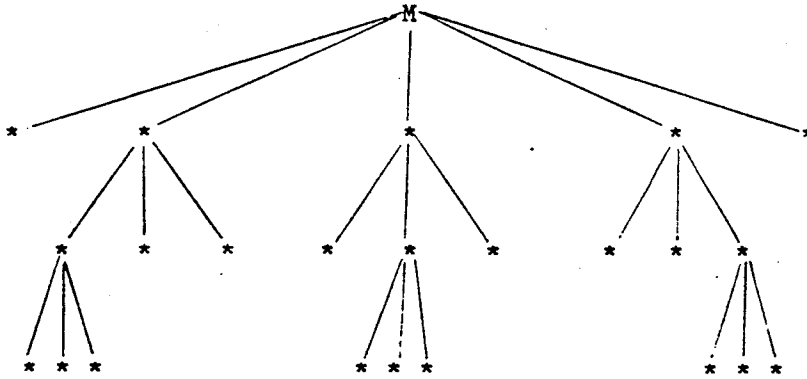


Figure 5-1: Search Tree

Here M represents the initial model, the asterisks immediately below it represent models that add one directed edge or correlated error to M, the next level of asterisks represent models that add two edges or correlated errors to M, and so on. The idea is that one first searches for the best additions at the first level below M, and ignores all additions at that level that are not among the best. One then considers all daughters of each of the best first level additions, and sees which among them are best, and so on. This sort of procedure is known as a **breadth-first search**. Notice that after the search through each level is completed, if the search is to continue then every daughter of the best nodes must be generated and evaluated. That means that to carry out the search we must have rapid ways to evaluate the effects of every possible addition to any model.

The breadth-first search strategy can be used with the TETRAD program, and we illustrate it in several cases in later chapters.

5.5. A SKETCH OF THE TETRAD PROGRAM

5.5.1. TETRAD's Output

The TETRAD program operates in two modes. One mode provides the user with information to conduct a breadth first search, one edge at a time, as described in the previous section. The other, more automatic mode, conducts a search that does not necessarily proceed one edge at a time, but which may consider several simultaneous additions to an initial model. The two modes can be used together.

TETRAD requires the user to provide the computer with a directed graph, representing an initial causal model, and a set of variances and covariances, or correlations, for the measured variables in the model. In addition the user can set a significance level to be used in testing constraints. With that input, the user can obtain any or all of the following information:

1. A list of all tetrad equations and vanishing partial correlations among measured variables that pass a statistical test at a level of significance specified by the user, as well as the p value of each constraint.
2. A list of all vanishing tetrad differences and vanishing partial correlations implied by the model given to the program, together with the p values of these constraints for the sample data provided.
3. A count of how many tetrad equations and vanishing partial correlations hold at a specified significance level, of how many of these are explained, i.e. implied, by the model given to the program, of how many constraints are implied by the model but do not hold (I-H), and of how many that hold are not implied (H-I).
4. The sum over all tetrad equations implied by the model, of (absolute values of) the tetrad differences found in the sample (**TTR value**), and the sum, over all vanishing partial correlations implied by the model, of (absolute values of) the partial correlations found in the data (**TPR value**).
5. Two tables, the **Rttr chart** and **Rtpr chart**, that list, for each possible directed edge or correlated error that may be added, the effects of adding that edge or correlated error to the initial model. The information includes I(H-I), D(I-H), Rttr (or Rptr), and Pi (see section 5.5.6 below).
6. For each possible addition of an edge or correlated error, a list of the tetrad equations or vanishing partials that are implied by the initial model but are no longer implied if the addition is made, and the probabilities of the corresponding residuals found in the sample (on the hypothesis that the differences vanish in the population). This option is called the "Compare" function.
7. A crude index of fit, the **Pi** value, for any model that is an elaboration of a basemodel specified by the user (see section 5.5.6 below).
8. Sets of suggested elaborations to a simple initial model, found by TETRAD's automatic search procedure.

We illustrate how these pieces of information can be used to find good causal models in section 5.5.8 below. First, however, some further explanation is needed of the information computed by the program.

5.5.2. Tetrad Residuals

For full statistical models with specified values for all parameters, we measure the difference between a correlation value implied by the model and the observed correlation by the *residual*. The residual is a function of the difference between a predicted correlation value and the observed correlation. The degree to which a tetrad equation implied by a model agrees with observed correlations can be measured in a similar way, but its calculation does not require that the model be estimated.

Consider for example the simple model from a previous chapter:

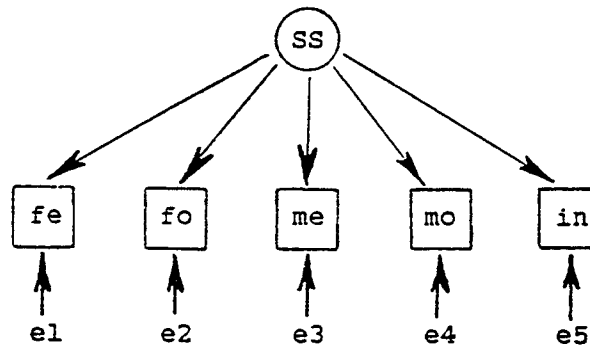


Figure 5-2: Directed Graph Representation

Consider a tetrad equation that is entailed by the model in Figure 5-2. The model implies that

$$\rho_{fo,me}\rho_{fe,mo} = \rho_{fo,mo}\rho_{me,fe}$$

If these variables stand for father's education, father's occupation, mother's education, mother's occupation, and family income for members of some population, we do not expect that sample correlations will exactly satisfy this equation, even if the model is correct. Suppose the observed correlations among these four variables, estimated from the sample, are

$$\sigma_{fo,me} = .433$$

$$\sigma_{fe,mo} = .381$$

$$\sigma_{fo,mo} = .521$$

$$\sigma_{me,fe} = .394$$

The tetrad equation implied by the model can be written equivalently as a vanishing tetrad difference:

$$\rho_{fo,me}\rho_{fe,mo} - \rho_{fo,mo}\rho_{me,fe} = 0$$

How much the observation differs from the implication can be measured by the positive square root of

$$(\sigma_{fo,me}\sigma_{fe,mo} - \sigma_{fo,mo}\sigma_{me,fe})^2$$

With the numbers just given, this quantity, which we call the **tetrad residual** is

$$((.433*.381 - .521*.391)^2)^{1/2} = .0387$$

One measure of how well a full statistical model with specified values for all of its parameters fits the data is the sum of the residuals of all correlations or covariances. Similarly, one measure of how well a causal model fits the data is the sum of the tetrad residuals of all the tetrad equations the model implies. As we said above, this quantity is the **Total Tetrad Residual** or **TTR**.

5.5.3. Partial Correlation Residuals

In the same way, we can consider the vanishing partial correlations that a model strongly implies, and compare them with the actual partial correlations found in the sample data. For each vanishing partial implied by a model, we call the corresponding absolute value of the sample partial correlation the **partial correlation residual** or **PR**. Just as with the sum of the tetrad residuals, we can sum the PR values for each partial correlation that must vanish in the population if the causal model is correct. We abbreviate this sum as **TPR**.

The TTR and TPR values are the heuristic measures we use for how well Thurstone's

Principle is satisfied by a model. The lower these values are, the better the Principle is satisfied.

5.5.4. The Statistical Procedures

Testing for Vanishing Tetrad Differences

For each foursome of measured variables in a data set, there are three possible tetrad equations which may, or may not, hold. For n variables the number of possible tetrad equations is therefore three times the number of ways of choosing 4 from n . The TETRAD program tests for each possible tetrad equation. The test is based on Wishart's formula for the variance of the sampling distribution of a tetrad difference. Consider four variables $x_1 - x_4$. Let D be the determinant of the sample variance-covariance matrix for the foursome of variables such that the (i,j) member of D is the covariance of x_i and x_j . Let D_{12} be the upper left corner two dimensional sub-determinant of D , and D_{34} the bottom right corner sub-determinant. N is the sample size. The formula for the sampling variance of the tetrad difference:

$$T = \text{Cov}(x_1, x_3)\text{Cov}(x_2, x_4) - \text{Cov}(x_1, x_4)\text{Cov}(x_2, x_3)$$

is given by:

$$(N-1)(N-2) \text{VAR}(T) = (N+1) D_{12} D_{34} - D$$

Other cases are obtained by permutation of indices. The distribution of the tetrad difference for correlations follows the same formula. The TETRAD program converts covariance tetrad differences to correlation tetrad differences to facilitate comparison of the tetrad residuals for different foursomes of variables.

TETRAD uses Wishart's formula to perform an asymptotic test for each possible tetrad equation. As the sample size grows without bound, the distribution of a tetrad difference converges in probability to a normal distribution (see Anderson [2]). The program compares the tetrad difference found in the data with a normal distribution having variance given by Wishart's formula and having mean zero.

Aside from its asymptotic character, there are several disadvantages to this test. The test is performed on each tetrad equation as though it were independent of the others. For many reasons, this assumption is almost never correct. The result is that in Monte Carlo studies, for example, we find that true tetrad equations are rejected by the test more frequently than the significance level would suggest. One might view the set of tetrad equations as a collection of *contrasts* (see Scheffe' , [86]), and seek simultaneous confidence intervals for all of the tetrad equations implied by a model as well as a test of the entire set of tetrad equations so implied. Such an approach poses a new and difficult statistical problem for each model, however, and we know of no general method for solving it or for computing such confidence intervals. In this case it is better to do what can be done, even if some error is introduced thereby. We regard the procedure as an example of Simon's "satisficing."

Another disadvantage of the tetrad difference statistic is that it is not very powerful when the correlations are small. The product of two small numbers is small, and the difference of such products (when the signs are the same) is smaller still. In Monte Carlo studies we have found that the statistic tends to be too generous in accepting tetrad equations when the sample sizes are small (below 300, for example) and the correlations are less than .3.

Testing for Vanishing Partial Correlations

For partial correlations, there is a well-known sampling distribution which can be transformed to a normal distribution using Fisher's z statistic (see Anderson, [2]). TETRAD tests for vanishing partial correlations using this statistic.

As with vanishing tetrad differences, the vanishing partial correlations will not generally be independent, and in testing all of the vanishing partial correlations implied by a model, the proportion of partials that vanish in the population but fail the statistical test may differ from the significance value of the test. We accept this limitation for the same reasons as in the case of the test for vanishing tetrad differences.

It should be emphasized that the hypotheses TETRAD tests are not the usual hypotheses of a "null effect." When testing null hypotheses, an investigator is often interested in rejecting the hypotheses, and thereby establishing that some effect of interest is significant. (The weakness of such tests has been discussed by Meehl , [76].) In such cases, increasing the significance level makes it more likely that the null hypothesis will be rejected, and therefore *weakens* the test of the hypothesis of interest. In contrast, when testing the hypothesis that a tetrad difference vanishes, one is usually interested in the case in which the hypothesis is *not* rejected. Increasing the significance level makes it more likely that the

hypothesis of a vanishing tetrad difference will be rejected, and therefore *strengthens* the test of this hypothesis. We will often test hypotheses at very high significance levels.

5.5.5. Determining the Constraints Implied by a Model

Given a data set and an initial model, the TETRAD program will compute all of the tetrad equations implied by the model. It will do the same for vanishing partial correlations, and it will report the sample tetrad difference and sample partial correlation for each constraint that is entailed by the model. The procedure is rigorous, and there is no restriction of any kind on the directed graph of the causal model. The graphs can have cycles, directed edges from measured to latent variables, whatever.

The procedures the program uses to compute the implied vanishing partial correlations are simply applications of the theorems in Chapter 4. The algorithms are described in more detail, and their adequacy proven, in Chapter 10. The procedures used to compute the vanishing tetrad differences implied by a model are not, however, simple applications of the theorems in Chapter 4. In models with cycles, the number of terms in a coordinating path effect can be infinite. The TETRAD procedure depends on the fact, proved in Chapter 10 that in all models, cyclic or not, the implied tetrad equations can be computed from the *open treks* alone.

5.5.6. Indices of Fit

Many indices of fit have been proposed in the literature on structural equation modelling. Bentler and Bonett proposed a "normed fit index" [5]. Let M_0 be a model that can be obtained from any of the models to be compared by fixing parameters. M_0 is thus more restrictive than any of the models to be compared. For example, M_0 might be the "null" model that fixes all structural parameters at zero. Let F be any fitting statistic one pleases. Let F_0 be the value of F for M_0 estimated from a sample covariance matrix S . In general, F_0 will be worse (i.e., larger) than the F value for any of the models to be compared. F_0 provides a standard, or unit, for how bad a model can be, and each of the models to be compared can be measured for fit by a fraction of F_0 . For any model M_t , Bentler and Bonett propose the index of fit:

$$\Delta_{0t} = (F_0 - F_t)/F_0$$

The index takes values between 0 and 1, and bigger is better, other things being equal.

The point of the normed index of fit is easily seen with a least squares loss function for F . In that case F is the sum of squares of residuals. F_0 is just the sum of squares of S , and Δ_{0t} is just unity, minus the squared sum of residuals of M_t as a fraction of the squared sum of residuals of S . The analogy with the notion of "percentage of variance explained" in analysis of variance is suggestive. A model can give a test statistic with a very low probability, even though it explains a very large proportion of the the sum of squares of S .

Nothing about the normed fit index requires that the function F be one for which a sampling distribution and appropriate statistical tests are available. Thus a normed fit index can be calculated for the sum of tetrad residuals. A natural choice for M_0 is a model that implies every possible tetrad equation among the measured variables. A single factor model in which each measured variable is an indicator of one latent factor, such that there are no directed or undirected edges connecting measured variables or their error terms, is a model that implies every possible tetrad equation among the measured variables. A model represented by a completely disconnected graph also implies every possible tetrad equation and every possible vanishing partial correlation equation.

TETRAD provides all of the information necessary to calculate a normed fit index using tetrad residuals.

James , et al. [48] have objected that the normed fit index does not take account of the number of free parameters in a model. They propose instead the index:

$$\Pi_{ok} = (d_k/d_0) [(F_0 - F_k)/F_0]$$

where d_k , d_0 are the degrees of freedom of the models M_k , M_0 respectively. The objection is not entirely fair, since Bentler and Bonett do not claim that normed fit ratios are by themselves a sufficient basis for preference among models, any more than the probability of a chi square statistic is such a basis. But James, et al., are surely correct that there is more to good explanation than simply reducing the proportion of the residuals unaccounted for, and that parsimony is part of what makes for a good explanation.

These indices have a certain arbitrariness, for two reasons. First, any sufficiently constrained model can be taken for M_0 , and different choices will lead to different values for the indices. The model that specifies that all variables are independent is a fairly

natural choice for M_0 . Second there is no natural or privileged measure of how much extra residual is worth tolerating for the gain of a degree of freedom. The Π index assumes such a measure, but it is fairly arbitrary. The Π index is in principle objectionable since it gives the worst possible value (zero) to any model which specifies that all variables are statistically independent, even when the data are in fact generated by such a structure. In practice, we are rarely interested in such models, and the objection does not much matter. For all the arbitrariness, it seems a substantial improvement to prefer models with greater Π values than to base preferences solely on likelihoods, or on Bentler and Bonnett's index of fit.¹⁸

TETRAD uses a surrogate for Π . The P_i value calculated by the program is based on the total tetrad residual, TTR, rather than on any usual fitting statistic. Otherwise, the principles are the same. The choice of a model M_0 in the TETRAD program is equally arbitrary, and we let the user make it. TETRAD's default for M_0 is the model with one common latent factor for all measured variables. This model is chosen because it implies every possible tetrad equation, and therefore no model can have a higher TTR value. The user is free to choose any other model for M_0 . The program must be given the TTR value for M_0 and it must be able to calculate the degrees of freedom of M_0 . For TETRAD to calculate the degrees of freedom it must know the number of edges in M_0 and the number of latent variables that are completely exogenous. No P_i value is calculated using partial correlations.

There is no simple connection between the residuals matrix and the sum of the tetrad residuals, although as the sum of squared residuals approaches zero, the sum of tetrad residuals must also vanish. For a fixed sample, the sum of tetrad residuals, TTR, can be changed only by altering the model and the set of constraints implied. Changes in the model can in principle reduce the TTR without reducing the sum of squares of residuals, although typically when one statistic is reduced so will the other one be. TETRAD uses the tetrad residuals rather than any of the usual fitting statistics because the TTR is a rapidly computable indicator that is minimally dependent on distribution assumptions.

5.5.7. TETRAD's Automatic Search Procedure

The search for alternative models using the TETRAD's Rtrr chart can be rather prolonged, and can require a lot of work on the part of the user. TETRAD has an automatic search procedure that can considerably shorten the search time required. The procedure is limited to models with latent variables.

¹⁸ Alternative fit indices are computed by the LISREL program, although the LISREL VI manual offers no argument for the particular indices chosen.

The automatic search procedure does not search for additional edges or correlated errors. Instead it searches for *treks* to be added to the initial model. It does not search one trek at a time, but instead considers the effects of the simultaneous addition of a *set* of treks. The output of the search is a *set of suggested trek additions* to the original model. The program may suggest several alternative sets of trek additions. The user must then use his or her judgement to decide how many of these treks to add, and to decide how to realize them as a combination of directed edges or correlated errors. The Rtrr chart can be of considerable help in this regard.

The automatic search procedure actually carries out repeated searches, as many times as the user requests. The suggested treks depend on which tetrad equations are judged to hold in the population, and that depends on the significance value chosen. In general, increasing values of the significance levels will result in the judgement that fewer tetrad equations hold, and that in turn will lead the search procedure to suggest more treks, or to add more sets of suggested trek additions.

How the Automatic Search is Conducted

TETRAD's search for suggested sets of trek additions carries out a strategy that insists on maximizing Spearman's principle, and consistent with that restriction, tries to maximize Thurstone's principle. For any set of tetrad equations judged to hold in the population, and any initial model, M , there are sets of trek additions to M that will imply every correct tetrad equation implied by M , and insofar as is possible consistent with that restriction, will not imply any tetrad equations that do not hold in the population. TETRAD's "Suggested Sets" are just these sets of trek additions to the initial model.

The TETRAD procedure for determining the sets of treks with the properties described is heuristic, not rigorous. In rare cases the program will suggest a trek that in fact prevents the implication of a tetrad equation that is implied by the initial model and that holds at the specified significance level.

Simple initial models usually imply a lot of tetrad equations, some of which hold at a given significance level and some of which do not. If everything implied holds, then no addition to the initial model can be made that prevents the implication of false tetrad constraints. As the significance level is increased, fewer and fewer tetrad equations will hold, and so more of the constraints implied by the initial model will be judged false, and larger numbers of trek additions are possible. The program increases the significance level from 0.0 until some tetrad equations implied by the initial model do not hold. Then the program finds all of the sets of simultaneous trek additions that will not prevent the implication of

any constraint that holds, but will prevent the implication (insofar as possible) of constraints that do not hold. The significance level is then increased again until some tetrad equations implied by the initial model and holding at the previous significance level no longer hold, and the search is conducted again. This process is repeated for as many iterations as the user requests. Two or three significance levels are usually enough.

A more detailed description of TETRAD's search algorithm is given in chapter 10 of this book.

Limitations of the Procedure

The Suggested Sets of trek additions are really intended for the analysis of multiple indicators, and may go badly astray for other kinds of initial models. The procedure has several requirements, and the conditions should be borne in mind when using the program.

- The procedure is heuristic, not rigorous.
- The initial model must be of a simple form. It must not have cycles, and it must not have directed edges from measured variables to other measured variables or from measured variables to latent variables. Every measured variable must depend on one and only one latent variable. Every latent variable must be connected (but not necessarily connected by a trek) with every other latent variable. We call the graphs of models meeting these conditions **skeletal**.
- The procedure assumes that any treks to be introduced between latent and measured variables are introduced as directed edges from the latent variable to the measured variable. In the output, suggested treks between latent variables and measured variables are denoted by directed edges, for example by "T → x1."

TETRAD will provide suggested sets of trek additions for models that are not skeletal, but it will give a warning that the model is improper.

5.5.8. An Example

To briefly illustrate how TETRAD's output is used, we describe a case that will be examined in much greater detail in Chapter 8. Recall the summary of TETRAD's output, given above:

1. A list of all tetrad equations and vanishing partial correlations among measured variables that **pass a statistical test** at a level of significance specified by

- the user, as well as the p value of each constraint.
2. A list of all vanishing tetrad differences and vanishing partial correlations implied by the model given to the program, together with the p values of these constraints for the sample data provided.
 3. A count of how many tetrad equations and vanishing partial correlations hold at a specified significance level, of how many of these are explained, i.e. implied, by the model given to the program, of how many constraints are implied by the model but do not hold (I-H), and of how many that hold are not implied (H-I).
 4. The sum over all tetrad equations implied by the model, of (absolute values of) the tetrad differences found in the sample (TTR value), and the sum, over all vanishing partial correlations implied by the model, of (absolute values of) the partial correlations found in the data (TPR value).
 5. Two tables, the Rtrr chart and Rptr chart, that list, for each possible directed edge or correlated error that may be added, the effects of adding that edge or correlated error to the initial model. The information includes I(H-I), D(I-H), Rtrr (or Rptr), and Pi (see section 5.5.6 below).
 6. For each possible addition of an edge or correlated error, a list of the tetrad equations or vanishing partials that are implied by the initial model but are no longer implied if the addition is made, and the probabilities of the corresponding residuals found in the sample (on the hypothesis that the differences vanish in the population). This option is called the "Compare" function.
 7. A crude index of fit, the Pi value, for any model that is an elaboration of a basemodel specified by the user.
 8. Sets of suggested elaborations to a simple initial model, found by TETRAD's automatic search procedure.

We will show each element of TETRAD's output in turn. The data are from M. Kohn's National Survey, reported in his *Class and Conformity*, [58]. They concern the answers to five questions on a much larger questionnaire. A factor analysis suggested that the responses to these questions are due to a single common factor, which Kohn took to be an authoritarian-conservative personality trait. 3101 white, non-farm males were sampled.

The initial measurement model is pictured in Figure 5-3. Here q1-q5 stand for the indicators and e1-e5 their disturbance terms.

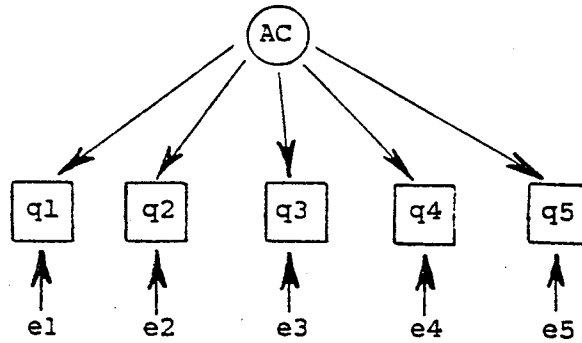


Figure 5-3: Initial AC Measurement Model

In the studies described by Kohn, the subjects were presented with a number of statements, and they were to indicate how strongly they agreed or disagreed. The questions were scored on a five category scale from "strongly disagree" to "strongly agree."

We will begin with the one factor model, and search for the best elaborations of it. The elaborations may be either directed edges, indicating that the answers given to some questions have direct effects on the answers given to other questions, or they may be correlated errors. First we give the initial model and the data to TETRAD and we ask for information concerning how well the model satisfies Spearman's and Thurstone's principles. The first chart below lists every tetrad equation that could possibly hold in the data, specifies the absolute value of the tetrad difference for that equation in the sample data, informs us whether the equation is implied by the model, informs us whether it passes a statistical test at an unusual significance level, .5, and gives us the p value of the tetrad difference in the sample on the hypothesis that the tetrad difference vanishes in the population. The reason for the significance level chosen will be explained in Chapter 8.

The graph analyzed in this example is:

AC→q1 AC→q2 AC→q3 AC→q4 AC→q5

The significance level is: 0.5000

The sample size is: 3101

Tetrad Equation	Residual	Impld.	Held	P(diff.)
q1 q2, q3 q4 = q1 q3, q2 q4	0.0068	y		0.3185
q1 q2, q4 q3 = q1 q4, q2 q3	0.0019	y	y	0.8099
q1 q3, q4 q2 = q1 q4, q3 q2	0.0087	y		0.2241
q1 q2, q3 q5 = q1 q3, q2 q5	0.0255	y		0.0000
q1 q2, q5 q3 = q1 q5, q2 q3	0.0295	y		0.0000
q1 q3, q5 q2 = q1 q5, q3 q2	0.0039	y	y	0.5163
q1 q2, q4 q5 = q1 q4, q2 q5	0.0160	y		0.0464
q1 q2, q5 q4 = q1 q5, q2 q4	0.0007	y	y	0.9193
q1 q4, q5 q2 = q1 q5, q4 q2	0.0167	y		0.0285
q1 q3, q4 q5 = q1 q4, q3 q5	0.0382	y		0.0000
q1 q3, q5 q4 = q1 q5, q3 q4	0.0051	y		0.3811
q1 q4, q5 q3 = q1 q5, q4 q3	0.0331	y		0.0000
q2 q3, q4 q5 = q2 q4, q3 q5	0.0204	y		0.0016
q2 q3, q5 q4 = q2 q5, q3 q4	0.0070	y		0.2063
q2 q4, q5 q3 = q2 q5, q4 q3	0.0134	y		0.0418

Figure 5-4: Statistical Information for the AC Skeleton

In TETRAD's output, we abbreviate the expression:

$$\rho_{i,j} * \rho_{k,l} = \rho_{i,k} * \rho_{j,l}$$

with

$$i j, k l = i k, j l$$

This kind of table covers items 1 and 2 in the list of TETRAD's output. The table shows, at a glance, that all possible tetrad equations are implied by the model, and that only three of them hold at a significance level of .5.

To have TETRAD automatically determine how well this model satisfies Spearman's and

Thurstone's principles, we ask for standard output and get the following information, which comprises items 3 and 4 in the list of output elements.

The graph analyzed in this example is:

AC->q1 AC->q2 AC->q3 AC->q4 AC->q5

The significance level is: 0.5000

The sample size is: 3101

The Total Tetrad Residual, or TTR, is: 0.22699

The Pi value for the model is: 0.000

I-H, the total number of equations implied by the model that do not hold in the data, is: 12

H-I, the total number of equations that hold in the data but are not implied by the model, is: 0

The number of equations that hold in the data is 3

The number of equations explained by the model is 3

The Total Partial Residual, or TPR, is: 0.00000

The Pi value for the model is: 0.000

PI-PH, the total number of equations implied by the model that do not hold in the data, is: 0

PH-PI, the total number of equations that hold in the data but are not implied by the model, is: 0

The number of equations that hold in the data is 0

The number of equations explained by the model is 0

Figure 5-5: Standard Results

The measures of how this model satisfies Spearman's principle for tetrad constraints, the number of equations explained and the (H-I) value, are perfect. This is to be expected of a model which implies every possible equation. It also trivially satisfies Spearman's principle for partial correlation constraints, as there are no vanishing partial correlations in the data at this significance level. Because of this we will not bother requesting any of TETRAD's output that concerns partial correlation constraints. Adding elaborations to a model that implies no partial correlation constraints cannot make it imply them.

The model does not satisfy Thurstone's principle for every tetrad constraint. (I-H) is 12,

which means that there are 12 equations the model implies which do not hold at a significance level of .5. The other measure of how well a model satisfies Thurstone's principle is the TTR. For this initial model it is .2269. In itself this number tells us little. What is important is how much of this TTR can be eliminated by elaborations of the model.

Now we turn to the task of finding good elaborations to this model. The next table, called the Rtrr Chart, gives us a list of each possible addition of a directed edge or correlated error. For each such addition, we are given the Rtrr value, i.e., how much each elaboration to the initial model reduces the TTR value of the initial model. We are also given $D(I-H)$, i.e., how many equations that are implied by the initial model but do not hold (at the .5 significance level) are no longer implied if the addition is made. We are further told $I(H-I)$, the difference between the initial model and the modified model with regard to how many equations hold in the data at the specified significance level but are not implied.

Edge	Rttr	D(I-H)	I(H-I)
q1-> q2	0.080	4	2
q2-> q1	0.080	4	2
q1 C q2	0.080	4	2
q1-> q3	0.088	5	1
q3-> q1	0.088	5	1
q1 C q3	0.088	5	1
q1-> q4	0.115	5	1
q4-> q1	0.115	5	1
q1 C q4	0.115	5	1
q1-> q5	0.089	4	2
q5-> q1	0.089	4	2
q1 C q5	0.089	4	2
q2-> q3	0.071	4	2
q3-> q2	0.071	4	2
q2 C q3	0.071	4	2
q2-> q4	0.067	5	1
q4-> q2	0.067	5	1
q2 C q4	0.067	5	1
q2-> q5	0.083	5	1
q5-> q2	0.083	5	1
q2 C q5	0.083	5	1
q3-> q4	0.067	5	1
q4-> q3	0.067	5	1
q3 C q4	0.067	5	1
q3-> q5	0.160	6	0
q5-> q3	0.160	6	0
q3 C q5	0.160	6	0
q4-> q5	0.087	5	1
q5-> q4	0.087	5	1
q4 C q5	0.087	5	1

Figure 5-6: Rttr Chart For AC Skeleton at .5

Recall that the best elaborations are those that have the *highest* Rttr value, the *highest* D(I-H) value, and the *lowest* I(H-I) value.

The Rttr chart suggests that either q3->q5, q5->q3, or q3 C q5 are the best single elaborations to the initial model. Any one of them does more than any other elaboration to reduce the TTR value, and is thus best by our heuristic for Thurstone's Principle. Further, none of these elaborations will prevent the model from implying any tetrad equation that

holds at the specified significance level and was implied by the initial model, i.e. their $I(H-I) = 0$. No alternative modifications do better. Thus these elaborations are also best by Spearman's Principle.

Notice that edges and correlated errors involving the pair $q1-q4$ seem second best. *One cannot conclude from this fact that directed edges or correlated errors connecting $q1-q4$ should be introduced at the second level of search.* Once $q3 \rightarrow q5$, $q5 \rightarrow q3$, or $q3 \text{ C } q5$ has been added to the initial model, the $q1-q4$ elaborations may no longer be the best at the second level. To find the best further additions, we would consider the initial model with $q3 \rightarrow q5$ added, the initial model with $q5 \rightarrow q3$ added, and the initial model with $q3 \text{ C } q5$ added. We would run all of these models through TETRAD again, asking for the Rtrr Charts. Here we will only show an Rtrr chart for the initial model with $q3 \rightarrow q5$ added.

The graph analyzed in this example is:

q3→q5 AC→q1 AC→q2 AC→q3 AC→q4 AC→q5

The significance level is: 0.5000

The sample size is: 3101

Edge	Rttr	D(I-H)	I(H-I)	Pi
q1→ q2	0.025	2	2	0.490
q2→ q1	0.025	2	2	0.490
q1 C q2	0.025	2	2	0.490
q1→ q3	0.042	4	2	0.534
q3→ q1	0.042	4	2	0.534
q1 C q3	0.025	3	1	0.488
q1→ q4	0.043	3	1	0.538
q4→ q1	0.043	3	1	0.538
q1 C q4	0.043	3	1	0.538
q1→ q5	0.026	2	2	0.493
q5→ q1	0.042	4	2	0.534
q1 C q5	0.026	2	2	0.493
q2→ q3	0.054	4	2	0.567
q3→ q2	0.054	4	2	0.567
q2 C q3	0.022	2	2	0.480
q2→ q4	0.033	3	1	0.510
q4→ q2	0.033	3	1	0.510
q2 C q4	0.033	3	1	0.510
q2→ q5	0.044	3	1	0.538
q5→ q2	0.054	4	2	0.567
q2 C q5	0.044	3	1	0.538
q3→ q4	0.038	4	2	0.522
q4→ q3	0.038	4	2	0.522
q3 C q4	0.021	3	1	0.478
q4→ q5	0.029	3	1	0.499
q5→ q4	0.038	4	2	0.522
q4 C q5	0.029	3	1	0.499

The Total Tetrad Residual, or TTR, is: 0.06700

The Pi value for the model is: 0.56348

Figure 5-7: Rttr Chart For AC Skeleton + q3→q5

Notice that after q3→q5 has been added, q1-q4 is no longer the elaboration that most reduces the TTR. Edges q5→q2, q2→q3, and q3→q2 are now the best according to

Thurstone's principle. According to Spearman's principle, however, $q1 \rightarrow q4$ is better than any of these edges. Its $I(H-I)$ value is only 1 while the other edges have an $I(H-I)$ value of 2. To see exactly what effect adding either of these edges has, we can use TETRAD's compare option, which is item 6 in the list of TETRAD's output. We input the skeleton + $q3 \rightarrow q5$ and ask for compare first on the edge $q1 \rightarrow q4$ and then on the edge $q5 \rightarrow q2$.

The graph analyzed in this example is:

$q3 \rightarrow q5$ AC- \rightarrow $q1$ AC- \rightarrow $q2$ AC- \rightarrow $q3$ AC- \rightarrow $q4$ AC- \rightarrow $q5$

Edge added: $q1 \rightarrow q4$

Tetrad Equation	Residual	Impld.	Held	P(diff.)
$q1 \ q2, \ q4 \ q3 = \ q1 \ q4, \ q2 \ q3$	0.0019		y	0.8099
$q1 \ q3, \ q4 \ q2 = \ q1 \ q4, \ q3 \ q2$	0.0087			0.2241
$q1 \ q4, \ q5 \ q2 = \ q1 \ q5, \ q4 \ q2$	0.0167			0.0285
$q1 \ q2, \ q4 \ q5 = \ q1 \ q4, \ q2 \ q5$	0.0160			0.0464

Rttr: 0.043 D(I-H): 3 D(H-I): 1

The graph analyzed in this example is:

$q3 \rightarrow q5$ AC- \rightarrow $q1$ AC- \rightarrow $q2$ AC- \rightarrow $q3$ AC- \rightarrow $q4$ AC- \rightarrow $q5$

Edge added: $q5 \rightarrow q2$

Tetrad Equation	Residual	Impld.	Held	P(diff.)
$q1 \ q2, \ q4 \ q3 = \ q1 \ q4, \ q2 \ q3$	0.0019		y	0.8099
$q1 \ q3, \ q4 \ q2 = \ q1 \ q4, \ q3 \ q2$	0.0087			0.2241
$q1 \ q3, \ q5 \ q2 = \ q1 \ q5, \ q3 \ q2$	0.0039		y	0.5163
$q1 \ q4, \ q5 \ q2 = \ q1 \ q5, \ q4 \ q2$	0.0167			0.0285
$q1 \ q2, \ q4 \ q5 = \ q1 \ q4, \ q2 \ q5$	0.0160			0.0464
$q2 \ q3, \ q5 \ q4 = \ q2 \ q5, \ q3 \ q4$	0.0070			0.2063

Rttr: 0.054 D(I-H): 4 D(H-I): 2

Figure 5-8: Compare Edges $q1 \rightarrow q4$ and $q5 \rightarrow q2$

From these charts we can see that the elaboration $q5 \rightarrow q2$ defeats a superset of the equations that $q1 \rightarrow q4$ defeats. The two extra equations that $q5 \rightarrow q2$ defeats are:

Tetrad Equation	Residual	Impld.	Held	P(diff.)
$q1\ q3, q5\ q2 = q1\ q5, q3\ q2$	0.0039		y	0.5163
$q2\ q3, q5\ q4 = q2\ q5, q3\ q4$	0.0070			0.2063

It is difficult to decide whether defeating these two equations is desirable. One of them holds fairly closely, so defeating it would violate Spearman's principle. On the other hand not defeating the equation that has $p = .2063$ would violate Thurstone's principle (at a significance level of .5). The decision should be made according to how the user weights the two principles.

Item 7 in the list of output elements is Pi. Pi can give a rough indicator of how to balance Thurstone's principle with simplicity. The more we elaborate a model the fewer constraints it implies and thus the more it satisfies Thurstone's principle. More complicated models have fewer degrees of freedom, however, and a model's performance on a statistical test depends on its number of degrees of freedom. At a certain point elaborations help Thurstone's principle but hurt a model's performance on a statistical test. Our Pi measure attempts to give the user some idea of when this point has been reached. Look at Figure 5-5 again. The model input will serve as the basemodel. Pi values for a basemodel are undefined, even though we output 0 for their value. In Figure 5-7 the model input is the skeleton plus the edge $q3 \rightarrow q5$. The Pi value for this model is .56348. What concerns us is whether the Pi value of further elaborations will be comparable to .56348 or higher. In the Rtrr chart shown in that Figure Pi values are listed for each candidate elaboration. These are the Pi values of the model input plus the elaboration shown in the Rtrr chart. The edge $q5 \rightarrow q2$, for example, has a Pi value of .567 in the Rtrr chart. This indicates that the skeleton + $q3 \rightarrow q5$ + $q5 \rightarrow q2$ has a Pi value of .567, which is higher than the skeleton + $q3 \rightarrow q5$. This suggests that $q5 \rightarrow q2$ is certainly a worthwhile elaboration. The Pi measure is only a heuristic guide, and we have found cases where adding an edge that decreases the Pi value still brings up the p value of the model.

In this case we would continue to search for a third elaboration. Continuing in this way we could form a search tree. A full search tree using the Rtrr chart can become quite large, even if one uses substantive assumptions and strong structural criteria to prune the tree at each level. The second level of a medium sized model can easily contain as many as twenty nodes, which means that to obtain the next level, the TETRAD program would have to be run twenty times. Since a run of the program can take several minutes, this search would be a tedious business if the investigator had to stay at the computer, running case after case. This task can be avoided by running TETRAD in **batch mode**. In batch mode the

user simply creates an input file for each case, and a command file that calls the TETRAD program, gives the program the appropriate commands, and names the input files and output files to be used. Batch files for a score of cases can be set up in a few minutes.

Finally, we illustrate TETRAD's automatic search for suggested elaborations on Kohn's initial model. One command to the TETRAD program produces the following information:

The graph analyzed in this example is:

AC→q1 AC→q2 AC→q3 AC→q4 AC→q5

The significance level is: 0.0500

The sample size is: 3101

Sets of suggested treks at significance level = 0.0418

{q3-q5 }

Sets of suggested treks at significance level = 0.5163

{q1-q3 q2-q5 q3-q5 }

Sets of suggested treks at significance level = 0.8099

{q1-q4 q2-q5 q3-q4 q3-q5 }
 {q1-q3 q2-q5 q3-q4 q3-q5 }
 {q1-q4 q2-q3 q3-q4 q3-q5 }
 {q1-q3 q2-q3 q2-q5 q3-q4 }
 {q1-q4 q2-q3 q2-q5 q3-q4 }
 {q2-q3 q2-q5 q3-q4 q3-q5 }
 {q1-q3 q2-q3 q2-q5 q3-q5 }
 {q1-q3 q1-q4 q2-q3 q3-q5 }
 {q1-q3 q1-q4 q2-q5 q3-q5 }
 {q1-q3 q1-q4 q3-q4 q3-q5 }
 {q1-q3 q1-q4 q2-q3 q2-q5 }
 {q1-q3 q1-q4 q2-q3 q3-q4 }

The Total Tetrad Residual, or TTR, is: 0.22699

I-H, the total number of equations implied by the model that do not hold in the data, is: 8

H-I, the total number of equations that hold in the data but are not implied by the model, is: 0

The number of equations that hold in the data is 7

The number of equations explained by the model is 7

Figure 5-9: TETRAD's Results on the AC Skeleton

There are a few simple heuristics we use with the suggested sets of trek additions.

1. We add the treks suggested at the lowest significance first, in all possible ways (consistent with substantive restrictions). Thus in this case we would add q5→q3 or q3→q5, just as we did using the Rtrr search strategy.

2. When a set of treks is added simultaneously, they may interact to form further treks that have not been suggested with the same set, or even at the same significance level. Thus $q5 \rightarrow q3$ and $q2 \rightarrow q5$ together form a trek connecting $q2 \text{ -- } q3$. In general realizations of treks that form further treks will further reduce the TTR value of the modified model, but they will also cause a loss in the ability to explain tetrad equations, and the user must judge whether the gain outweighs the loss. Our rule of thumb is that we prefer implementations of suggested treks at a given level which create further treks not suggested at that level only if all treks implied by a modification are suggested at the next significance level of suggested sets. Thus, for example, we would consider $q1 \rightarrow q3$, $q5 \rightarrow q3$, $q2 \rightarrow q5$, even though a $q2 \text{ -- } q3$ trek is created because the set

$$\{q1-q3 \ q2-q3 \ q2-q5 \ q3-q5 \}$$

is suggested at the next level of significance.

5.6. THE VALUE OF P VALUES

In applying TETRAD, we routinely assess the models found using full information maximum likelihood estimation with the LISREL or EQS programs, and associated statistical tests. We view the probabilities of test statistics, or p values as they are sometimes called, as *likelihoods*. Likelihoods are just probabilities conditional on hypotheses, and probabilities can be compared. The probability of the chi square statistic for some model and data set is a measure both of fit and simplicity, for it takes into consideration both the residuals and the number of free parameters used in producing those residuals. Hypotheses can be compared by the ratio of the likelihoods they give to the same test statistic on the same data. Of course, that is not all there is, or should be, to the comparison of competing hypotheses, but it is a consideration, and we will use it repeatedly.

The use of statistical tests has a decision theoretic foundation in Neyman-Pearson theory, and the most devoted adherents of the tradition reject any theory comparisons that are not framed as hypothesis tests. They would, in particular, reject the comparison of alternative models in terms of likelihoods, or simplicity, or explanatory power. The Neyman-Pearson decision theory is rarely strictly applied, but when it is it does not permit anything but the *rejection* of theories. The acceptance of a theory, however tentatively, cannot be justified on Neyman-Pearson grounds, save in those rare cases in which every alternative theory has been rejected. Nor does the decision theory provide any judgements about which of several *false* theories is the *better approximation*. But in all of science, including social

science, the principal job is not to reject false theories, but to find the better approximations, and to reject the worse approximations.

In practice, in most applications of linear modeling, we can be fairly confident that every model we will consider is false. The almost universal experience is that as sample sizes increase, testable linear models fail powerful statistical tests. The task is not to find a *true* linear model, but to find the models that are the best linear approximations to the truth, and to assess whether that approximation is good enough to serve in practice. In such contexts, there is little point to decision procedures whose function is to reject false theories.

There is, to our knowledge, no formal account of the connection between properties of sample statistics, on the one hand, and the approximate truth of models, on the other. In the absence of such an account, we use the probabilities of chi square values, simplicity, and explanatory power to assess models, and where we can we test models by means of non-statistical predictions.

5.7. TETRAD AND OTHER SEARCH PROCEDURES

Several procedures have been proposed for searching for alternative linear causal models, given an initial model. Anything that works is to be applauded, but we believe TETRAD has some considerable advantages. Alternative procedures include the following:

1. Examination of Residuals.

One strategy is to locate the correlation with the largest residual, and to free a structural parameter associated with a causal connection between the measured variables thus correlated. Costner and Schoenberg [17] have shown that the strategy is, as one might expect, often misleading. For example, if a model implies several tetrad equations that are not even approximately satisfied in the data, the addition of an edge or set of edges connecting variables whose residual correlations are not maximal may be the most efficient way to modify the model so that the false constraints are not implied. A closely related strategy examines normed residuals, in which the residuals are assumed to be normally distributed and are represented in units of the variance of the distribution.

2. Partial derivatives of a fitting statistic.

Byron [13] and in a modified form Sorbom [98], have proposed a strategy that is now

embodied in the LISREL programs. The strategy is to treat a statistic as a function, for given sample correlations, of the fixed parameters of the model, and to consider the partial derivatives of the function with respect to each parameter. For example, the partial derivatives may be evaluated at the maximum likelihood estimate of the parameters. The parameter with the largest squared partial derivative (and appropriate second derivative) is freed. The LISREL VI program contains "modification indices", which are the ratios of these first derivatives and corresponding second derivatives, and the program will, if desired, free parameters one after another according to the size of their modification indices. In other terms, LISREL VI conducts a beam search through the space of alternative models. Sorbom noted that the strategy may be misleading because it is applied sequentially, one parameter at a time, and also because a large change in a parameter with a small partial derivative may produce better fit than a small change in a parameter with a larger partial derivative. In Chapter 8 we have seen examples, such as the Wheaton data on alienation, in which the strategy does seem to pass by the best models. Costner and Herting [18] have explored the limitations of the strategy, and suggested that it is not reliable for detecting connections between indicators of a common latent variable. They also point out that it may lead to errors if it is *not* applied one parameter at a time.

3. Testing sub-models.

For multiple indicator models, Costner and Schoenberg described a mixed strategy for specifying revisions. Part of the strategy amounted to eliminating tetrad residuals by examining two factor submodels, each with two indicators, and adding directed edges (and appropriate free parameters) to such submodels when they fail chi square tests. The modified submodels do not entail a tetrad equation implied by the unmodified submodel, and thus the modification eliminates a tetrad residual. Costner and Schoenberg did not propose carrying out an analogous procedure, however, for other kinds of submodels.

Any strategy for finding good alternatives to an initial model is bound to be heuristic, and that means that it is bound to fail in some circumstances. There are practical and structural reasons, however, why the nested model search conducted by the LISREL program is especially inadequate, and since the program is widely used, they deserve to be mentioned.

Consider the tree of increasingly less constrained models described in an earlier section. The LISREL search strategy starts with an initial, presumably simple, model as does the TETRAD strategy. But LISREL determines at each stage a *unique* model at the next level. That means that the LISREL procedure considers only a single path through the tree of alternative models. Searches of this kind are called **beam searches** in the artificial intelligence literature. For the reasons mentioned above, the decision at each stage as to

which is the best parameter to free may be unreliable, and this fragility is compounded by the beam search. A second difficulty with the LISREL procedure is that it gives *no weight* to preserving the ability of the initial model to explain constraints satisfied by the data. Spearman's Principle has no role in the search. A third difficulty has to do with stopping the search. The LISREL search strategy stops when the chi square difference between a model obtained at stage n of the search and the next model, obtained at stage $n + 1$, is not statistically significant. The trouble is that models which are elaborations of the $n+1$ st model, and which are never considered by the LISREL search, may be very significant, and may have chi square differences (when compared with the n th stage model) that are very significant. For example, we will later describe a study of alienation for which, starting with an initial model that has a chi square p value of zero, LISREL finds an elaborated model with a chi square p value of .335. The LISREL procedure adds two free parameters to the initial model, and then stops. But with TETRAD we find a further elaboration of the LISREL model, freeing two more parameters, that has a chi square p value of .99. Together, the two further modifications found with TETRAD make a statistically significant difference in the fit. A fourth, practical, difficulty with the LISREL procedure is that the LISREL formalism derives from factor analysis, and because of those origins the formalism does not permit direct causal relations between measured variables. That means that if one represents a causal model in the most straightforward way in LISREL, the search procedure will not discriminate between directed edges and correlated errors between measured variables. The result is that one cannot detect the direction of omitted directed edges between measured variables, as one can with TETRAD. This limitation can be overcome in principle by the artifice of using dummy theoretical variables for measured variables, but in practice the strategem makes the representation of the models more cumbersome, and may often fail to be used. Using the TETRAD program it is straightforward to find many plausible models that exhibit much better fit than those produced by LISREL's beam search.

6. WHAT TETRAD CAN DO

6.1. INTRODUCTION

In the preceding chapters we have claimed a lot for the TETRAD program, and some readers may well be sceptical. We claim that researchers with TETRAD will do better science than researchers without it. That is a difficult claim to substantiate, and a lot of the remainder of the book is devoted to mustering evidence for it. To give the reader a sense of what can be done with the program, we present in this chapter a number of problems that are easily solved with TETRAD's help. The answers to the problems, and the methods used to find those answers, are given in later chapters. Using the demonstration program that accompanies this book, the reader can repeat the same procedures we have used. The answers to problems 1, 2, 3 and 6 are given in Chapter 8; the answers to questions 4 and 5 are given in Chapter 7.

For now, however, we ask the reader simply to consider the following questions: Can the problems be solved *without* using the TETRAD program? We urge the reader to use whatever techniques he or she may favor to attempt to answer the problems, and see if those techniques do as well as the TETRAD program.

In real science the answers are not in the back of the book, and sometimes there are no answers. Some data sets simply cannot be modelled adequately by any remotely plausible linear theory, and some data sets can be accounted for about equally well by a vast collection of alternative linear causal models.

Heuristic search can be of help in these circumstances too, although none of the problems of this chapter show how. On real data sets, TETRAD sometimes yields results that suggest that no linear model will adequately explain the data. In other cases, the program gives results that suggest that too many linear models will explain the data.

6.2. ALIENATION

To do this problem one must use LISREL or EQS or some other program that provides maximum likelihood estimates and chi square tests.

Joreskog and Sorbom [46] discuss two models of data on alienation taken from a study done by Wheaton et al [101]. The models are used to illustrate the operation of the LISREL programs.

Their first model, pictured in Figure 1-0, has a chi square probability (with 6 degrees of freedom) of zero.

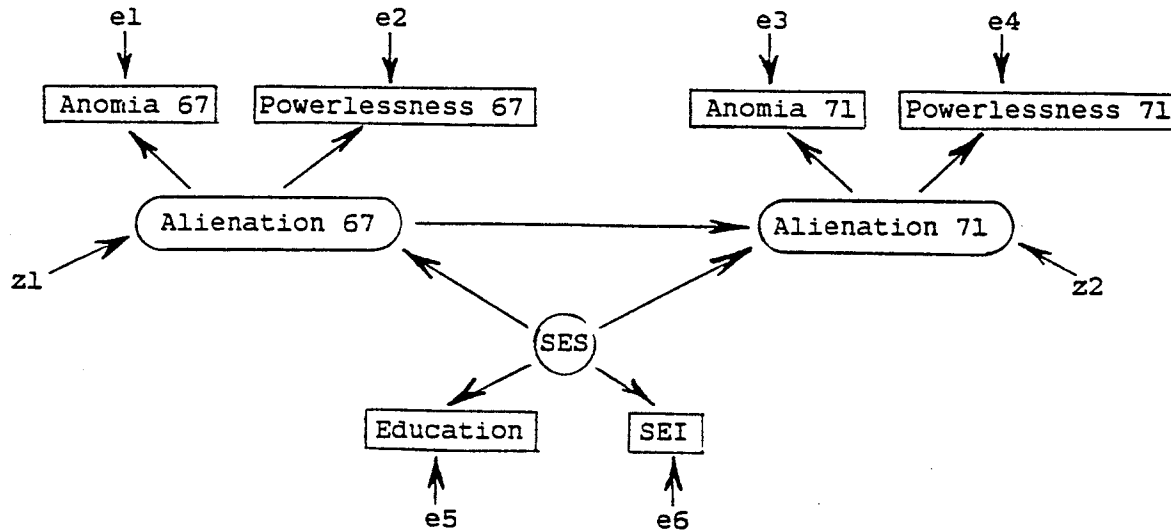


Figure 6-1: Alienation: Original Model

Joreskog and Sorbom amend this model by freeing two parameters and arrive at the model, shown in Figure 6-2, with 4 degrees of freedom and $p = .335$, $n = 932$.

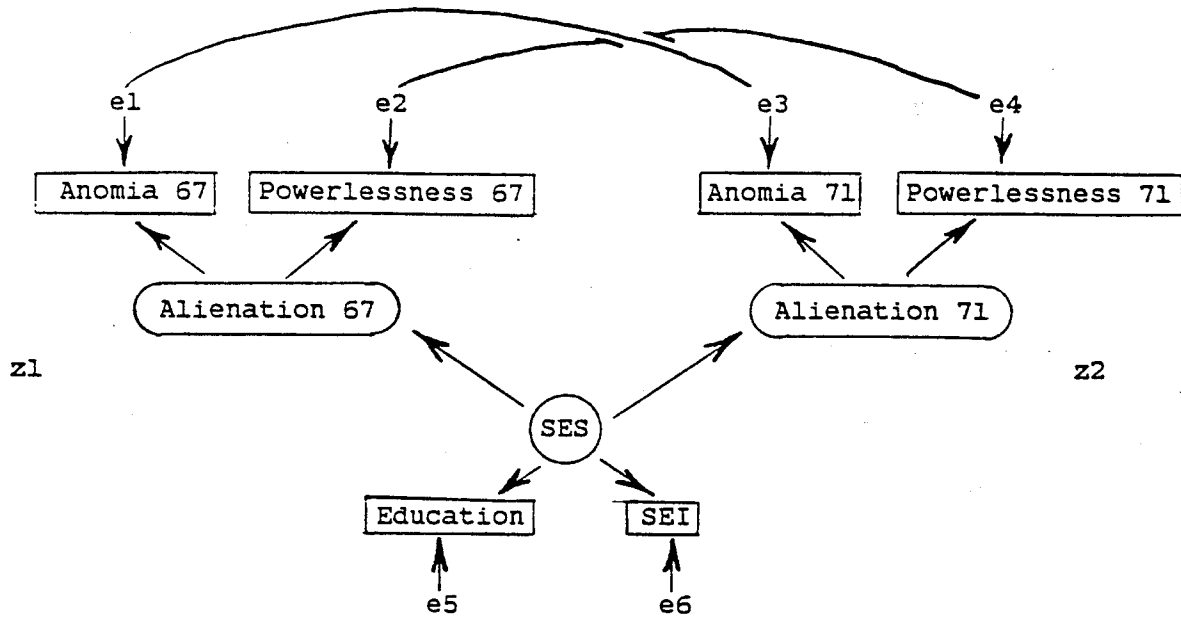


Figure 6-2: Alienation: Amended Model

The variance-covariance matrix for this problem, with obvious abbreviations, is taken from the User's Manual to LISREL IV [46].

	A6	P6	A7	P7	ED	SE
A6	11.834					
P6	6.947	9.634				
A7	6.819	5.091	12.532			
P7	4.783	5.028	7.495	9.986		
ED	-3.839	-3.889	-3.841	-3.625	9.610	
SE	-21.899	-18.831	-21.748	-18.775	35.522	450.288

Figure 6-3: Correlations for the Alienation Study

Here are two problems:

Problem 1: Find a model that frees 3 parameters in the original model (and thus has three degrees of freedom) and gives a chi square statistic with probability greater than .8.

Problem 2: Find a model, distinct from that in problem 1, that frees 3 parameters in the the original model and gives a chi square statistic with probability greater than .9.

Since the solutions to problems 1 and 2 are both about as plausible as the initial model or the LISREL revision, one might ponder the risks involved in failing to discover these

alternative theories.

6.3. A PROBLEM USING SIMULATED DATA

In the studies in this section and the next, we generated data using SYSTAT BASIC. We produced data for the exogenous variables with a pseudo random number generator distributed normally with mean 0 and variance 1. All other variables are linear combinations of their respective immediate ancestors in the directed graph of the model. The linear coefficients were also chosen at random, although their values for each particular model are non-stochastic constants.

6.3.1. Study 1

For the first problem we generated correlational data from a model that is an elaboration of Figure 6-4.

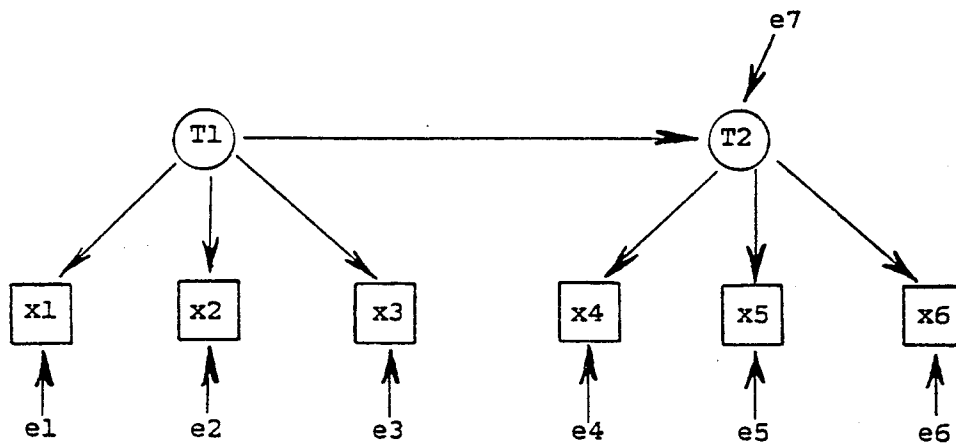


Figure 6-4:

T1 and T2 are to be interpreted as unmeasured latent variables, and e1-e7 as unmeasured error terms. T1 and e1-e7 are distributed normally with mean 0 and variance 1. x1-x7 are our measured variables. The sample size is 5000. The correlation matrix is shown in Figure 6-5.

	x1	x2	x3	x4	x5
x2	0.25120				
x3	0.49724	0.30191			
x4	0.36036	0.20448	0.42635		
x5	0.32524	0.51905	0.38348	0.48400	
x6	0.23871	0.12966	0.27900	0.44256	0.47144

Figure 6-5: Correlation Data For Study 1. $n = 5000$

Two edges that were in the model that generated the data have been removed from the figure shown above. These did not include any edges from measured variables to latent variables. They are either directed edges from latent to measured variables, directed edges from measured to measured variables, or correlated error terms. The true model has no more than one edge between each pair of variables.

Problem 3: Find the two missing edges. You do not have to give a single answer. The fewer alternative answers however, the better, as long as the correct answer is among the alternatives.

The space of possible answers to this problem contains 2,190 members.

TETRAD correctly picks out one edge uniquely. It narrows the choices for the second edge to six; running these six models through the EQS program, and picking those that have significant p values for their chi square statistics leaves three models, including the correct answer.

6.4. CAUSAL ORDER FROM CORRELATIONS

6.4.1. Monte Carlo Data: I

The three models pictured in Figure 6-6 all imply different overidentifying constraints. Specifically, they all imply different sets of tetrad equations.

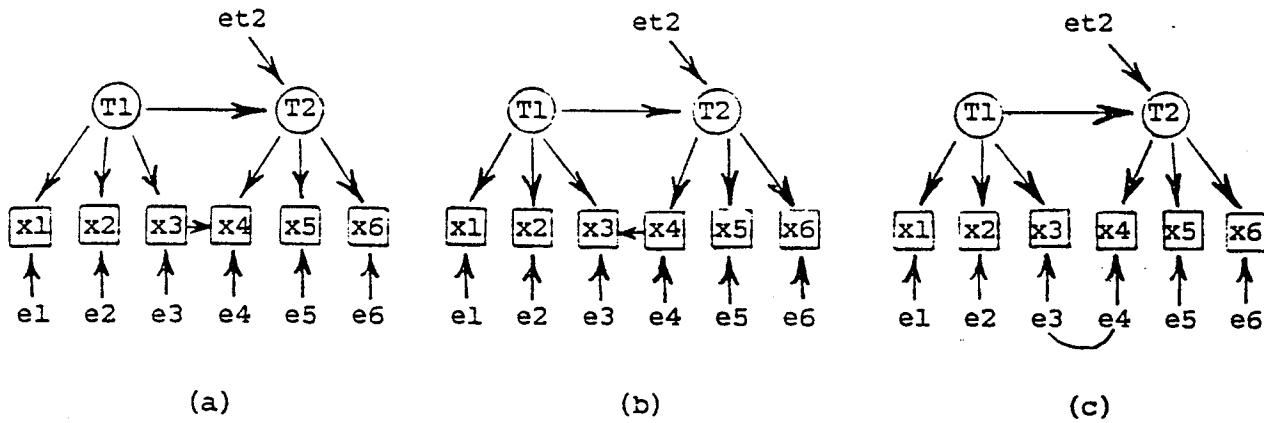


Figure 6-6:

We generated five data sets for this problem. For each of the five data sets, we randomly selected a model from the above three and then generated data with $n = 2000$ in the same way we have described above. The problem is as follows.

Problem 4: For each of the five data sets below, determine whether model (a), model (b), or model (c) generated the data.

The chance of randomly choosing the correct sequence of models is 1 in 243. In five minutes a TETRAD user identified the sequence perfectly.

Data Set 1

	x1	x2	x3	x4	x5
x2	0.70895				
x3	0.70653	0.61196			
x4	0.77571	0.67201	0.87201		
x5	0.75803	0.65612	0.63334	0.84425	
x6	0.69713	0.59854	0.58938	0.78916	0.87428

Data Set 2

	x1	x2	x3	x4	x5
x2	0.73218				
x3	0.71264	0.61605			
x4	0.65140	0.56910	0.88900		
x5	0.75321	0.65899	0.64722	0.82760	
x6	0.69263	0.60523	0.60106	0.76872	0.87026

Data Set 3

	x1	x2	x3	x4	x5
x2	0.72900				
x3	0.70424	0.61704			
x4	0.65653	0.57492	0.90401		
x5	0.76259	0.66237	0.66115	0.66115	
x6	0.71318	0.62563	0.63339	0.78092	0.87924

Data Set 4

	x1	x2	x3	x4	x5
x2	0.73119				
x3	0.77805	0.68057			
x4	0.68019	0.60167	0.92926		
x5	0.76382	0.67034	0.85745	0.84872	
x6	0.70363	0.62473	0.79043	0.79385	0.87863

Data Set 5

	x1	x2	x3	x4	x5
x2	0.74892				
x3	0.70336	0.62088			
x4	0.65350	0.56444	0.89529		
x5	0.75899	0.65678	0.63346	0.81258	
x6	0.71367	0.60912	0.58985	0.75083	0.86361

Figure 6-7: Data Sets For Monte Carlo I

6.4.2. Monte Carlo Data: II

Problem 4 was relatively easy because the only pair of variables with an ambiguous source of extra correlation was x_3 and x_4 . In this problem we remove that constraint. We generated five data sets from a random elaboration of the skeleton shown in Figure 6-4. The possible elaborations are not restricted to causal connections between x_3 and x_4 . Any indicator of T1 may be connected with (or have correlated errors with) any indicator of T2. Thus *one* of the following set of alternative elaborations of the model in Figure 6-4 comprises the possibilities for each data set.

```

{  x1 -> x4   x4 -> x1   e1 <--> e4
   x1 -> x5   x5 -> x1   e1 <--> e5
   x1 -> x6   x6 -> x1   e1 <--> e6
   x2 -> x4   x4 -> x2   e2 <--> e4
   x2 -> x5   x5 -> x2   e2 <--> e5
   x2 -> x6   x6 -> x2   e2 <--> e6
   x3 -> x4   x4 -> x3   e3 <--> e4
   x3 -> x5   x5 -> x3   e3 <--> e5
   x3 -> x6   x6 -> x3   e3 <--> e6 }

```

Figure 6-8: Possible Elaboration For Monte Carlo II

Problem 5: Identify the correct elaboration of the model in Figure 6-4 in each of the following five data sets.

The chance of choosing the correct sequence by chance is less than 1 in 14,000,000. A TETRAD user correctly identified every member in the sequence, and, exclusive of creating the data files, the entire process required only about five minutes.

Data Set 1

	x1	x2	x3	x4	x5
x2	0.70137				
x3	0.54959	0.58843			
x4	0.77796	0.54506	0.41529		
x5	0.76528	0.53461	0.41923	0.76241	
x6	0.92669	0.55317	0.43270	0.79971	0.77696

Data Set 2

	x1	x2	x3	x4	x5
x2	0.73537				
x3	0.56398	0.57400			
x4	0.52982	0.54909	0.40887		
x5	0.51391	0.54306	0.40730	0.77331	
x6	0.72118	0.89617	0.56778	0.75928	0.74619

Data Set 3

	x1	x2	x3	x4	x5
x2	0.71903				
x3	0.51791	0.55951			
x4	0.50208	0.53428	0.39059		
x5	0.72145	0.90772	0.54715	0.72531	
x6	0.51754	0.54350	0.36724	0.80703	0.73396

Data Set 4

	x1	x2	x3	x4	x5
x2	0.71301				
x3	0.57291	0.59103			
x4	0.78770	0.51065	0.40642		
x5	0.50273	0.52626	0.38593	0.72558	
x6	0.53287	0.56519	0.41803	0.76173	0.77435

Data Set 5

	x1	x2	x3	x4	x5
x2	0.70716				
x3	0.50878	0.56222			
x4	0.52308	0.50910	0.37597		
x5	0.50341	0.50189	0.38013	0.74525	
x6	0.48956	0.49444	0.77257	0.73827	0.73473

Figure 6-9: Data Sets For Monte Carlo II

6.4.3. Kohn's Study and Temporal Order Among Interview Questions

M. Kohn's *Class and Conformity* describes several large survey studies that investigate the relationships between social class, attitudes, and personality structure [58]. One section of his book explores the relations between social class and a variety of latent factors, one of which is an "Authoritarian-Conservatism" (hereafter abbreviated "AC ") personality trait.

The measured variables for AC are answers to interview questions, measured with a five category scale from "strongly disagree" to "strongly agree." Five of the questions are:

1. The most important thing to teach children is absolute obedience to their parents.
2. In this complicated world, the only way to know what to do is to rely on leaders and experts.
3. No decent man can respect a woman who had sexual relations before marriage.
4. Any good leader should be strict with people under him in order to gain their respect.
5. It's wrong to do things differently from the way our forefathers did.

The correlations among these five variables are shown in Figure 6-10.

	q1	q2	q3	q4	q5
q1	1.0				
q2	0.7161	1.0			
q3	0.4505	0.3356	1.0		
q4	0.7746	0.5012	0.3525	1.0	
q5	0.4263	0.3421	0.3149	0.3014	1.0

Figure 6-10: Correlation Matrix for the Authoritarian Trait Indicators

The initial measurement model is pictured in Fig. 6-11.

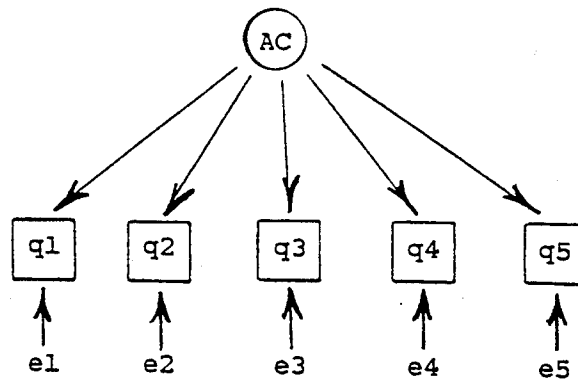


Figure 6-11: Initial AC Measurement Model

Note carefully that *the numbering of the questions is not necessarily the order in which the questions were asked on Kohn's questionnaire*. There are five questions, and the number of distinct possible orderings of them is $5!$ or 120.

Problem 6: Find the order in which the five questions were asked on Kohn's survey.

Again a unique answer is not required, but the smaller the set of alternatives the better. With TETRAD, the number of alternatives was reduced from 120 to 40 possibilities, and the true ordering of the questions was among the forty.

142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

7. USING TETRAD TO DETERMINE CAUSAL ORDER

7.1. INTRODUCTION

In Chapter 6 we posed a number of problems in which data were generated by a known linear model. The reader was given the data, the sample size, and *part* of the model used to generate the data. The problem was to find the remainder of the model. Here we are interested in exhibiting how TETRAD handles these problems.

In the first of these problems, we tell the user which pair of measured variables has a source of correlation¹⁹ in the model that generated the data but not in the model that we exhibit. The task is to determine whether this pair of variables have correlated error terms, or whether instead the variables are directly causally related. That is, *the user must distinguish correlation from causation on the basis of non-experimental correlation data and a known causal skeleton*. Further, if the pair of variables *is* judged to be directly causally connected in the model that generated the data, then we ask the user *to determine the direction of causation*.

In the second of these problems we impose the same demands as in the first, but we add one: the user is asked to determine *which pair of variables* has a source of correlation in the model that generated the data that is not shown in the model given to the user.

7.2. DISTINGUISHING CORRELATION FROM CAUSATION

The following problem has five parts. First, consider three models:

¹⁹we say "is connected "

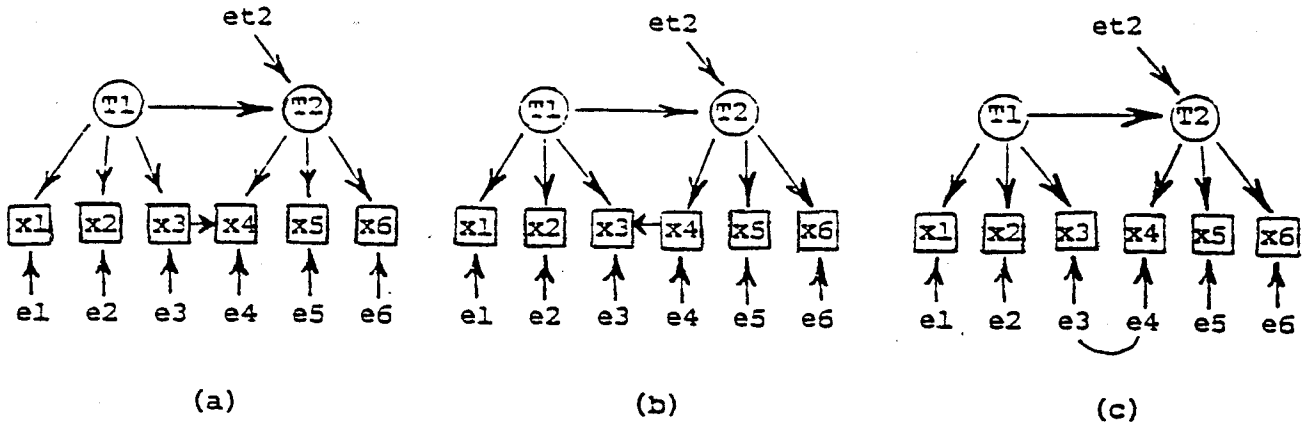


Figure 7-1:

Each of these models is an elaboration of the following *skeletal* model:

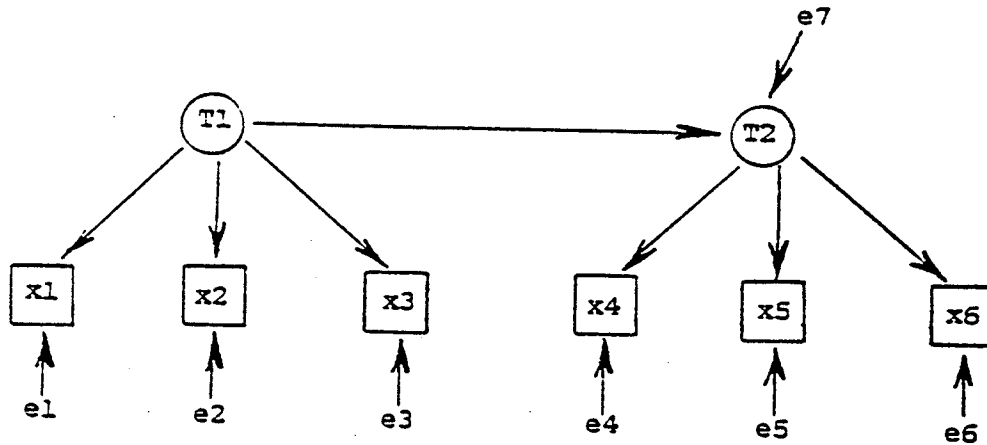


Figure 7-2: Skeletal Causal Model

We generated five data sets. For each data set, we randomly selected a model from the three elaborations of the skeletal model and then generated data with $n = 2000$. The problem was as follows:

Problem 4: For each of the five data sets below (see page 133), determine whether model (a), model (b), or model (c) generated the data.

The TETRAD strategy for solving these problems is easy. We give the common skeletal graph and the data set to TETRAD. We then ask for the Rtrr chart at a significance level

of .05. We choose this level because the sample size is large, and we know that all of the normality assumptions implicit in our statistical tests are strictly satisfied, and because the Suggested Sets feature of the program already suggests additions at this significance level. The Rtrr chart tells us how much each possible directed edge or correlated error will reduce the TTR value of the initial model, and gives us the value of $I(H-I)$ for each possible addition. $I(H-I)$, it will be recalled, is the increase in the number of equations that hold but are not implied.

For each data set, we choose the edge or correlated error with $I(H-I) = 0$ that has the highest Rtrr value. It should be emphasized that the results given below were obtained by a user who did not know the correct answers beforehand. The TETRAD output for the first data set is shown below. For the readers convenience, we have surrounded our choice of the missing edge or correlate errors with "*"s. For data sets 2 - 5 we simply list our choices.

Edge	Rtrr	D(I-H)	I(H-I)
x1→ x2	0.328	2	4
x2→ x1	0.328	2	4
x1 C x2	0.328	2	4
x1→ x3	0.166	1	5
x3→ x1	0.166	1	5
x1 C x3	0.166	1	5
x1→ x4	1.311	9	3
x4→ x1	0.662	5	7
x1 C x4	0.635	5	3
x1→ x5	0.628	4	8
x5→ x1	0.437	3	9
x1 C x5	0.282	2	6
x1→ x6	0.635	4	8
x6→ x1	0.432	3	9
x1 C x6	0.274	2	6
T2→ x1	0.170	1	5
x2→ x3	0.170	1	5
x3→ x2	0.170	1	5
x2 C x3	0.170	1	5
x2→ x4	1.263	9	3
x4→ x2	0.593	5	7
x2 C x4	0.569	5	3
x2→ x5	0.607	4	8
x5→ x2	0.405	3	9
x2 C x5	0.254	2	6

x2-> x6	0.609	4	8
x6-> x2	0.401	3	9
x2 C x6	0.244	2	6
T2-> x2	0.166	1	5
** x3-> x4	1.718	12	0 **
x4-> x3	1.440	8	4
x3 C x4	1.397	8	0
x3-> x5	0.791	5	7
x5-> x3	0.918	5	7
x3 C x5	0.619	3	5
x3-> x6	0.761	5	7
x6-> x3	0.881	5	7
x3 C x6	0.567	3	5
T2-> x3	0.328	2	4
x4-> x5	0.460	3	3
x5-> x4	0.460	3	3
x4 C x5	0.460	3	3
x4-> x6	0.435	3	3
x6-> x4	0.435	3	3
x4 C x6	0.435	3	3
T1-> x4	0.846	6	0
x5-> x6	0.846	6	0
x6-> x5	0.846	6	0
x5 C x6	0.846	6	0
T1-> x5	0.435	3	3
T1-> x6	0.460	3	3

Our guess is that the edge $x3 \rightarrow x4$ is missing from the skeleton.

We list all our guesses in Table 7-1 below.

Data Set	Missing Edge or Correlated Error
1	x3 C x4
2	x3 C x4
3	x3-->x4
4	x4-->x3
5	x3 C x4

Table 7-1: Our Guesses For Problem 4

In each case, the answer chosen is correct.

7.3. FINDING WHICH PAIR OF VARIABLES HAS AN EXTRA SOURCE OF CORRELATION

The problems so far were relatively easy because we know that the omitted edge or correlated error connected x3 and x4. Recall that the second set of Monte Carlo studies in Chapter 6 were more difficult in principle, because we were not told in advance which pair of variables were connected. But in practice, one problem is quite as easy as the other with the TETRAD program.

We generated five data sets from a random elaboration of the skeleton shown in Figure 7-2. The possible elaborations are not restricted to causal connections between x3 and x4. Any indicator of T1 may be causally connected with or have correlated errors with any indicator of T2. Thus the following set of alternative elaborations of the model in Figure 7-2 comprises the possibilities for each data set.

```

{  x1 -> x4   x4 -> x1   e1 <--> e4
   x1 -> x5   x5 -> x1   e1 <--> e5
   x1 -> x6   x6 -> x1   e1 <--> e6
   x2 -> x4   x4 -> x2   e2 <--> e4
   x2 -> x5   x5 -> x2   e2 <--> e5
   x2 -> x6   x6 -> x2   e2 <--> e6
   x3 -> x4   x4 -> x3   e3 <--> e4
   x3 -> x5   x5 -> x3   e3 <--> e5
   x3 -> x6   x6 -> x3   e3 <--> e6 }

```

Figure 7-3: Alternative Elaborations For Problem 5

Recall problem 5 in Chapter 6.

Problem 5: Identify the correct elaboration of the model in Figure 7-2 in each of the five data sets shown on page 139.

Our procedure is exactly as in problem 4. We give the TETRAD output for the first data set, and then we simply list our guess for each of data set.

Edge	Rttr	D(I-H)	I(H-I)
x1-> x2	0.573	3	3
x2-> x1	0.573	3	3
x1 C x2	0.573	3	3
x1-> x3	0.565	3	3
x3-> x1	0.565	3	3
x1 C x3	0.565	3	3
x1-> x4	0.593	5	7
x4-> x1	1.362	9	3
x1 C x4	0.567	5	3
x1-> x5	0.582	5	7
x5-> x1	1.362	9	3
x1 C x5	0.554	5	3
x1-> x6	0.952	8	4
** x6-> x1	1.569	12	0
x1 C x6	0.932	8	0
T2-> x1	1.120	6	0
x2-> x3	1.120	6	0
x3-> x2	1.120	6	0
x2 C x3	1.120	6	0
x2-> x4	0.391	3	9
x4-> x2	0.659	4	8
x2 C x4	0.261	2	6
x2-> x5	0.367	3	9
x5-> x2	0.655	4	8
x2 C x5	0.247	2	6
x2-> x6	0.612	5	7
x6-> x2	0.726	5	7
x2 C x6	0.403	3	5
T2-> x2	0.565	3	3
x3-> x4	0.380	3	9
x4-> x3	0.663	4	8
x3 C x4	0.265	2	6
x3-> x5	0.360	3	9

x5→ x3	0.664	4	8
x3 C x5	0.247	2	6
x3→ x6	0.571	5	7
x6→ x3	0.698	5	7
x3 C x6	0.367	3	5
T2→ x3	0.573	3	3
x4→ x5	0.217	2	4
x5→ x4	0.217	2	4
x4 C x5	0.217	2	4
x4→ x6	0.131	1	5
x6→ x4	0.131	1	5
x4 C x6	0.131	1	5
T1→ x4	0.136	1	5
x5→ x6	0.136	1	5
x6→ x5	0.136	1	5
x5 C x6	0.136	1	5
T1→ x5	0.131	1	5
T1→ x6	0.217	2	4

The missing edge is x6-->x1.

We list all our guesses for problem 5 below.

Data Set	Missing Edge or Correlated Error

1	x6-->x1
2	x2-->x6
3	x2-->x5
4	x1 C x4
5	x3 C x6

Table 7-2: Our Guesses For Problem 5

In each of these cases TETRAD was able to give us a unique, and correct answer because the modeling assumptions were strictly realized, the sample size was large, and only one additional edge or correlated error was required. In cases in which several edges or

correlated errors must be inferred from an initial model and data, the program will not usually fix upon a single alternative, but it will serve to dramatically reduce the possibilities. An example of this kind was given in Chapter 6, and it will be treated in Chapter 8.

8. CASE STUDIES

8.1. INTRODUCTION

This chapter describes the application of TETRAD to a number of cases. Most of the data sets we discuss are empirical rather than simulated, and have been considered by several other authors. Wheaton's study on alienation, for example, has been discussed by Joreskog and Sorbom in the LISREL manual and elsewhere, and by Bentler in the EQS manual; data from summer Head Start have been considered by a number of authors, and so have the data from Kohn's National Survey. The empirical cases have been deliberately chosen to permit the reader to compare the results found with, and without, TETRAD's aid. They, and the studies done with simulated data, are also intended to illustrate the use of search heuristics with the TETRAD program. All but one of the empirical data sets we discuss can be analyzed with the version of TETRAD on the disc accompanying this book.

The discussions of the procedures used in these cases are deliberately repetitive. We hope that repetition will promote more familiarity than contempt.

The case studies illustrate several important points:

- TETRAD can help find models for empirical data that are both plausible and that do well on a chi square test, models that have been overlooked by unaided workers.
- TETRAD can help search for alternative skeletal models. In particular it can help the user to decide whether or not latent variables ought to be introduced to account for the data.
- Assuming an initial model, TETRAD can help the user discriminate between correlated errors and direct connections between variables, and can even discriminate between alternative causal directions for direct connections.
- The program makes it easy to take advantage of the user's prior knowledge in the search for adequate models. In fact, effective use of the program requires such knowledge to eliminate hypotheses that do well by TETRAD's criteria but make no causal sense.
- TETRAD allows the user to carry out a much more thorough search for alternative models than can be done by hand, or with the aid of LISREL's beam search.

8.2. INDUSTRIAL AND POLITICAL DEVELOPMENT

8.2.1. Introduction

In 1973, H. Costner and R. Schoenberg proposed a method for identifying misspecifications in multiple indicator models. Their procedure is like the TETRAD procedure in several important respects. Costner and Schoenberg illustrated their technique with a model of data for industrial and political development.

They consider the model:

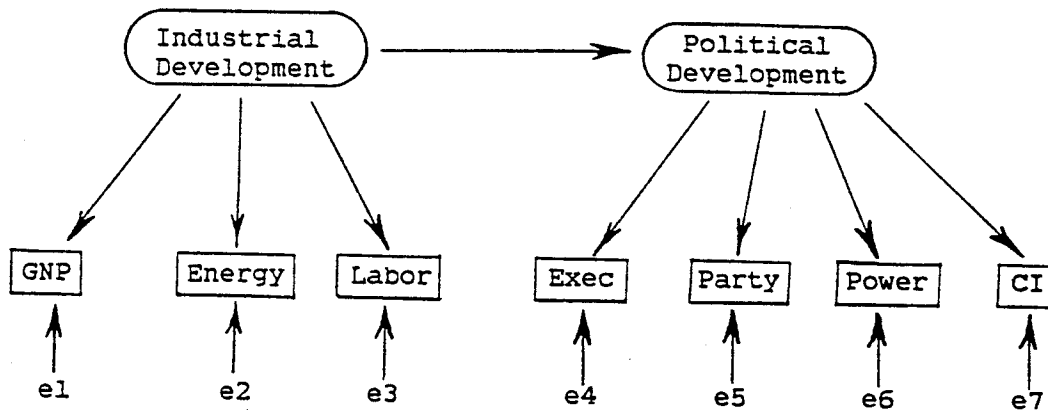


Figure 8-1: Costner and Schoenberg's Initial Model

In this model, GNP is gross national product, Energy is the logarithm of total energy consumption in megawatt hours per capita, Labor is a labor force diversification index, Exec is an index of executive functioning, Party is an index of political party organization, Power is an index of power diversification, and CI is the Cutright Index of political representation.

Using their revision procedure, Costner and Schoenberg arrive at the following modification of the initial model:

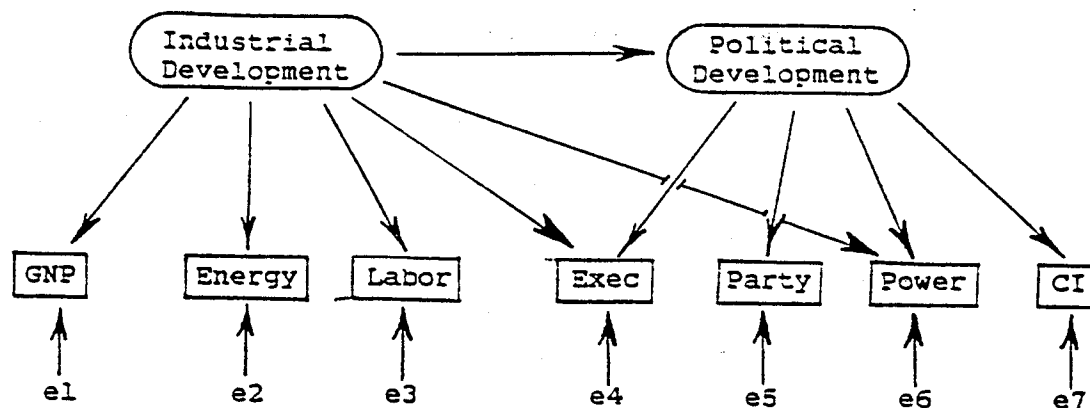


Figure 8-2: Costner and Schoenberg's Revised Model

The revised model has a chi square statistic with $p = .053$, and 11 degrees of freedom. The fit is marginal. One might want to know if there are good alternatives to this model or if the fit can be improved by adding further connections. To answer these questions we use TETRAD, starting with Costner and Schoenberg's initial causal model. We have included the Costner and Schoenberg data set on the TETRAD disk. The file name is cs.dat. We urge the user to follow the analysis below with TETRAD and then experiment with the program by trying other lines of search.

8.2.2. The TETRAD Analysis

We will do two searches, first with the Suggested Sets procedure, and then with the Rtrr chart. We begin with the skeletal causal model, shown in Fig. 8-3. The abbreviation scheme we use is shown below:

Industrial Development	=	ID
Political Development	=	PD
GNP	=	gp
Energy	=	en
Labor	=	la
Exec	=	ex
Party	=	pa
Power	=	po
Cutright Index	=	ci

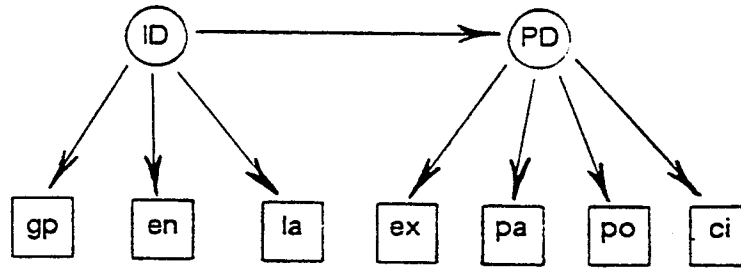


Figure 8-3: C & S Skeleton

We create a covariance file which we show in Fig. 8-4. The first line of a covariance file contains the sample size. Each remaining line contains a covariance (or correlation) or variance and then a variable pair. The same variable name is given twice in each line in which a variance occurs.

```

115
1.0 gp gp
1.0 en en
1.0 la la
1.0 ex ex
1.0 pa pa
1.0 po po
1.0 ci ci
0.95 gp en
0.83 gp la
0.66 gp ex
0.56 gp pa
0.45 gp po
0.67 gp ci
0.83 en la
0.70 en ex
0.54 en pa
0.38 en po
0.66 en ci
0.62 la ex
0.54 la pa
0.38 la po
0.61 la ci
0.47 ex pa
0.45 ex po
0.60 ex ci
0.64 pa po
0.64 pa ci
0.67 po ci
  
```

Figure 8-4: C & S Data File

We also create a graph file, which we show below. Each line of the graph file contains an edge in the graph. Latent variables are capitalized and measured variables are in lower case.

```
ID PD
ID gp
ID en
ID la
PD ex
PD pa
PD po
PD ci
```

Figure 8-5: Graph File For The Skeleton

Using the Sets of Suggested Treks

We proceed to use TETRAD's run format on this data file, requesting 2 units of suggested sets. TETRAD's output is shown below.

The graph analyzed in this example is:

```
ID->gp ID->en ID->la ID->PD PD->ex PD->pa PD->po PD->ci
```

The sample size is: 115

Sets of suggested treks at significance level = 0.5316

```
{ID->ex ID->po }
{ex-po pa-ci ID->po }
{ex-po ID->ex pa-ci }
```

Sets of suggested treks at significance level = 0.6027

```
{ex-po ID->ex pa-po pa-ci }
{ex-po ID->ex pa-ci po-ci }
{ex-po ID->ex pa-po po-ci }
{ex-po pa-po pa-ci po-ci }
{pa-po pa-ci po-ci ID->po }
{ex-po pa-po pa-ci ID->po }
{ex-po pa-ci po-ci ID->po }
{ex-po ID->ex pa-po ID->po }
{ID->ex pa-po pa-ci ID->po }
{ID->ex pa-po po-ci ID->po }
{ex-po ID->ex po-ci ID->po }
{ID->ex pa-ci po-ci ID->po }
```

The Total Tetrad Residual, or TTR, is: 5.29850

Figure 8-6: C & S Output

Notice that the first of the suggested sets is precisely the one that Costner and Schoenberg locate with their procedure. Adding those edges to the skeleton and asking for standard results, we can find the TTR of the revised model. Table 8-1 below shows the TTR, degrees of freedom, chi-square and its associated probability for the skeleton and the model Costner and Schoenberg suggest.²⁰

Notice the difference in the TTR of the skeleton and Costner and Schoenberg's revision. The TTR, recall, is just the sum of the tetrad residuals of all of the tetrad equations implied by the model.

²⁰We used EQS/PC to compute the chi-square and its associated probability. We fixed at 1: a coefficient for one indicator of each latent construct.

Edge Additions	TTR	DF ²¹	X ²	p(X ²)
None (skeleton)	5.298	13	52.1	<0.001
ID->po, ID->ex	1.149	11	19.5	0.053

Table 8-1: Statistical Results For the Models C & S Consider

The goodness of fit of a model is negatively correlated with the TTR. Adding edges to a model prevents the model from implying certain tetrad equations and the effect of adding an edge depends on which equations are no longer implied by the modified model. If the equations that are no longer implied have large tetrad residuals, the TTR will be substantially reduced. If those equations have a small residuals, explanatory power will be lost without much reduction in the TTR value. Some edges will reduce the TTR substantially but will also prevent the model from explaining equations that hold closely in the data. The user is free to weight these factors as he or she wishes.

In its sets of suggested trek additions, however, TETRAD puts priority on preventing the loss of explanatory power over reducing the TTR. For this case that weighting is for the most part too severe. ID->ex and ID->po are good suggestions, and they are also the edges occurring in the suggested sets which most reduce the TTR value. But there are a number of edges not occurring in the suggested sets that reduce the TTR even more and that lose little in explanatory power. This conclusion can be drawn immediately from an examination of the Rtrr chart for the original model.

Using the Rtrr Chart

The Rtrr chart provides powerful information in a general way. Given *any* linear model, it informs the user of the exact effect the addition of any single edge has on this model's TTR, its D(I-H), its D(I-H), and its Pi value. The chief disadvantage is that it gives this information only for *single* edge additions. Thus it is not generally appropriate to find two edges which look promising in an Rtrr chart and to conclude that both should be added. The two edges may interact in a way not anticipated, they may defeat the same equations and be redundant, and so forth.

The most effective way to use just the Rtrr chart is to perform a tree-like search, in

²¹Degrees of freedom.

which each node of the tree is a graph which results from the addition of a single edge to its "parent" graph. Start with a plausible initial model at the root of the tree, run TETRAD, and ask for suggested sets to find a significance level to use (see below). Then run TETRAD to obtain an Rtrr chart for the initial model at that significance level. From this chart pick a set of the best individual edges to add. Add these edges to the initial model one at a time to form the first level of the tree. Run TETRAD on each graph in the first level to obtain an Rtrr chart for each of them. Pick the best members from the Rtrr chart for each node and form a second level, etc. Any path down the tree gives a series of nested models. The only difficulty is in knowing at what point adding edges will no longer help improve a model's fit. TETRAD's P_i value may sometimes be useful in making this decision.

The program for Suggested Sets of trek additions finds its significance levels automatically. The Rtrr chart does not. The Rtrr chart requires the specification of a significance level, which is used to decide whether or not a tetrad equation "holds" in the population. The set of tetrad equations that "hold" is then used to discriminate among the explanatory powers of alternative one edge or one correlated error additions to the initial model. The default value of the significance level is the conventional .05, but generally users should select *unconventional* values for the significance level, especially when the sample size is not large. One is not making a real decision as to which tetrad equations literally hold in the population: in all likelihood, none of them do. One is instead trying to find those patterns or constraints that are well approximated. The most important thing in selecting a significance value is to find a level that will efficiently discriminate among alternative models.

Our recommendation is the following:

Run the Suggested Sets procedure before running the Rtrr chart. Find the significance levels at which sets of trek additions are first suggested. Choose that value, or higher values, as the significance level for the Rtrr chart.

In this case we use the significance level of the second unit of suggested sets of trek additions, .532. Asking for an Rtrr chart for the skeleton at a significance level of .532, we obtain the following results.

Edge	Rttr	D(I-H)	I(H-I)
(Not in a suggested set)			
gp-> ex	3.178	24	2
gp-> po	2.693	25	1
en-> ex	3.186	24	2
en-> po	2.697	25	1
la-> ex	3.111	22	4
la-> po	2.593	24	2
(In a suggested set)			
ex-> po	1.422	14	0
po-> ex	1.422	14	0
ex C po ²²	1.422	14	0
ID-> ex	2.920	18	0
pa-> po	2.003	13	1
po-> pa	2.003	13	1
pa C po	2.003	13	1
pa-> ci	1.553	14	0
ci-> pa	1.553	14	0
pa C ci	1.553	14	0
po-> ci	1.617	13	1
ci-> po	1.617	13	1
po C ci	1.617	13	1
ID-> po	2.407	18	0

The Total Tetrad Residual, or TTR, is: 5.29850

I-H, the total number of equations implied by the model that do not hold in the data, is: 56

H-I, the total number of equations that hold in the data but are not implied by the model, is: 1

The number of equations that hold in the data is 14

The number of equations explained by the model is 13

Figure 8-7: Rttr Chart At .532

²²The notation "ex C po" means that the error terms for ex and po are correlated.

We have included only those edges not in the suggested sets that seem especially promising. There are a number of edges that considerably reduce the TTR value without extracting a high cost in explanatory power, and many of the treks in the second unit of suggested sets of trek additions are not among the best of these. We expect that additions to the skeleton included in the suggested sets, with the exception of ID→po and ID→ex, will not substantially improve the fit of the model, and that expectation is confirmed by the findings summarized in the following table.

Model	(Additions to Skeleton)	p(X ²)
1)	ID→po, ID→ex	0.053
2)	ID→po, po→ex	0.037
3)	ID→po, po→ex, ci C po	0.049
4)	ID→ex, ex→po, pa C po, ci C po	0.037
5)	ID→ex, ID→po, pa→po	0.054
6)	ID→ex, ID→po, po C ci	0.056

Table 8-2: The Best Models From The Suggested Sets

These models are only marginally significant. We can do better by seeking revisions with the help of the Rtrr chart.

The Search

Adding edges to a model reduces the TTR but also reduces the degrees of freedom. After a certain number of additions, the reduction in the TTR is no longer large enough to offset the loss in simplicity and explanatory power. In adding two edges, Costner and Schoenberg have substantially lowered the TTR, but their revision still has a TTR of 1.149. Pi helps indicate if adding another edge might improve the fit still further. We illustrate a full tree like search below.

We start with the skeleton and request a full Rtrr chart with the significance level set at .532. The Pi values are calculated using the skeleton as a base model. Already screened results were shown above in Figure 8-7. ; We construct the following tree, shown below. Each edge and correlated error is shown along with the TTR and Pi value of the model that

results from its addition.

Root -----	Level 1 -----		
Initial Model	Edge	TTR	Pi
	ID->ex	2.38	.509
	ID->po	2.85	.419
	gp->ex	2.12	.554
	gp->po	2.61	.469
	en->ex	2.11	.555
	en->po	2.60	.470
	la->ex	2.19	.542
	la->po	2.71	.452

Figure 8-8: Tree With First Level Added

We now run TETRAD on each of the eight models occupying a node in the first level of the tree and request an Rtrr chart for each. We also ask TETRAD to give us Pi values for the models on the first level and to indicate in the Rtrr charts the Pi values of models that are elaborations of the first level. To illustrate, we show part of the Rtrr chart and standard results for the most promising model on level one, the skeleton plus the edge en->ex .

Edge	Rttr	D(I-H)	I(H-I)	Pi
gp-> ex	0.091	2	2	0.523
ID-> po	1.231	14	0	0.705
gp-> po	1.452	20	1	0.741
en-> po	1.426	19	1	0.737
la-> po	1.386	19	2	0.730
ex-> po	1.426	19	1	0.737

The Total Tetrad Residual, or TTR, is: 2.11220

The Pi value for the model is: 0.55510

I-H, the total number of equations implied by the model that do not hold in the data, is: 32

H-I, the total number of equations that hold in the data but are not implied by the model, is: 3

The number of equations that hold in the data is 14

The number of equations explained by the model is 11

Figure 8-9: Rttr Chart For Skeleton + en->ex

Notice that the Pi value for the model is .555, as the Rttr chart for the skeleton indicated it would be. We include the edge gp->ex in this chart to illustrate that it may be unwise to add both of two edges that do well in the same Rttr chart. Look again at Figure 8-7. Edges gp->ex and en->ex both look extremely promising. After we add en->ex, however, the Rttr chart shows that gp->ex is not a good further addition. En->ex and gp->ex are almost entirely redundant.

We illustrate how to form the second level of the tree by showing the level under en->ex, the most promising node.

Root -----	Level 1 -----			Level 2 -----		
	Edge	TTR	Pi			
	(ID->ex	2.38	.509)			
	(ID->po	2.85	.419)			
	(gp->ex	2.12	.554)			
	(gp->po	2.61	.469)	(gp-> po	.660	.741)
Initial Model	(en->ex	2.11	.555)	(ID-> po	.881	.705)
	(en->po	2.60	.470)	(en-> po	.686	.737)
	(la->ex	2.19	.542)	(la-> po	.726	.730)
	(la->po	2.71	.452)	(ex-> po	.686	.737)

Figure 8-10: Tree With Part of Second Level Added

If we add the edge gp->po the resulting model has a Pi value of .741. The Pi value is increasing rapidly, and the models with two additions still explain most of the equations that hold, so searching for a third addition seems certainly worthwhile. We add each of the five edges in turn and ask TETRAD for an Rtrr chart. Below we show part of our third level under the additions gp->po and en->po.

Root	Level 1			Level 2		Level 3	
-----	Edge	TTR	Pi				

	(ID->ex	2.38	.509)				
	(ID->po	2.85	.419)				
	(gp->ex	2.12	.554)			(pa->en .237 .735)	
	(gp->po	2.61	.469)	(gp-> po	.660 .741)	(ci->en .237 .735)	
Initial	(en->ex	2.11	.555)	(ID-> po	.881 .705)	(la->pa .206 .739)	
Model	(en->po	2.60	.470)	(la-> po	.726 .730)		
	(en->ex	2.11	.555)	(ex-> po	.686 .737)	(pa->en .297 .734)	
	(en->po	2.60	.470)	(en-> po	.686 .737)	(la->pa .252 .740)	
	(la->ex	2.19	.542)			(ci->la .301 .733)	
	(la->po	2.71	.452)				

Figure 8-11: Tree With Part of Third Level Added

The Pi values of the third level are leveling off and in some cases decreasing. This indicates that searching at the fourth level might be fruitless.

Proceeding in this way for the entire tree, we arrive at 20 distinct models that reduce the TTR to between approximately .2 and .3 and that add only three edges to the skeleton. We then run all twenty models through EQS/PC²³, getting parameter estimates and goodness of fit results. The results are in Table 8-3 below.

8.2.3. Goodness of Fit Results

²³See chapter 11 for help in running twenty models in batch format.

Additions to Skeleton	TTR	Eqs. Explained	$p(X^2)$
1) en->po, gp->ex, la->pa	.146	7	0.0551
2) en->po, gp->ex, en->pa	.220	6	0.1189
3) en->po, gp->ex, po->pa	.220	6	0.1002
4) en->po, la->pa, en->ex	.202	6	0.4792
5) en->po, la->pa, po->ex	.202	6	0.4085
6) en->po, la->ex, ex->pa	.223	5	0.1108
7) en->ex, la->po, pa->la	.198	4	0.0246
8) en->ex, la->po, ci->la	.198	4	0.0161
9) gp->ex, pa->po, en->pa	.220	6	<0.001
10) gp->ex, gp->po, la->pa	.220	6	<0.001
11) en->po, en->ex, gp->pa	.270	5	0.5752
12) en->po, en->ex, ci->la	.241	4	0.5198
13) en->po, la->ex, pa->en	.213	3	0.1234
14) ID->po, en->ex, ex->gp	.358	5	0.1222
15) ID->po, en->ex, pa->en	.328	5	0.1526
16) en->ex, pa->po, la->pa	.272	5	0.0438
17) en->ex, pa->po, en->pa	.286	6	0.0211
18) en->ex, pa->po, ex->pa	.286	6	0.0096
19) gp->ex, gp->po, la->pa	.209	6	<0.001
20) en->po, la->ex, la->pa	.222	6	0.0819

Table 8-3:

Note that in several cases there are models that are almost indistinguishable by TETRAD's measures but that perform quite differently on a chi-square test. There are four models that have large p values ($p(X^2) > .4$) and a number that are well above the .05 cutoff. The skeleton for this model implies higher order constraints, e.g. 50 pentad equations, that may be violated in the data and that are not considered by the TETRAD program. Of models that appear indistinguishable by TETRAD's measures but behave quite differently on chi square tests, some may imply false higher order constraints not implied by others.

Notice that all four models that do extremely well involve direct connections between en-po and en-ex. Also notice that the only models that have $p(X^2) < .001$ include neither of these connections. Of the two, en-po seems more important. No model that involves en-po is insignificant at the .05 level, while models that involve en-ex but not en-po have $.01 < p(X^2) < .05$.

Testing TETRAD's Suggestions About Causal Order

This series of models also offers support for a hypothesis concerning the causal order between en-po and en-ex. The Rtrr chart distinguishes three possible connections among a pair of variables. Notice that every model we located with the Rtrr tree search specifies the en-po connection as en→po, and the ex-en connection as en→ex.

To test TETRAD's prediction about this ordering we ran a number of variants of the ordering through EQS. We exhibit the results in Table 8-4 below.

Edge Additions To Skeleton	p(X^2)
1) en→po, en→ex	0.5504
2) po→en, en→ex	0.0336
3) en→po, ex→en	0.1434
4) po→en, ex→en	<.001
5) enCpo, enCex, exCpo	0.0015
6) T→en, T→ex, T→po	0.0015 ²⁴
7) ID→po, en→po, en→ex	0.6082
8) ID→po, po→en, ex→en	0.0477
9) ID→ex, en→po, en→ex	0.6004
10) ID→ex, po→en, en→ex	0.0931
11) ID→ex, ID→po, PD→en	0.2940

Table 8-4: Causal Order Comparisons

The results are unambiguous. In every case in which two models differ with respect to a causal ordering, the model that accords with TETRAD's suggestion does markedly better on a chi-square test.

Finally, we exhibit a table of all the models we located with TETRAD's help, their TTR value, the number of equations which hold at a significance level of .532 that the models explain, and the p value of the chi square statistic for each model. The models are ordered by the probability of their statistic.

²⁴ Here T is an extra latent variable introduced as a common cause of en, ex and po.

Model	TTR	Eqs. Explained ²⁵	p(X ²)
Skeleton	5.300	13	0.000
1) en->po, en->ex, ID->po	.686	10	0.6082
2) en->po, en->ex, ID->ex	.686	10	0.6004
3) en->po, en->ex, gp->pa	.270	5	0.5752
4) en->po, en->ex, pa->en	.246	4	0.5680
5) en->po, en->ex	.686	10	0.5504
6) en->po, en->ex, ci->la	.251	4	0.5198
7) en->po, en->ex, la->pa	.202	6	0.4792
8) en->po, la->pa, po->ex	.202	6	0.4085
9) ID->po, ID->ex, PD->en	.560	5	0.2940
10) ID->po, en->ex, pa->en	.328	5	0.1526
11) ex->en, en->po			0.1434
12) en->po, la->ex, pa->en	.213	3	0.1234
13) ID->po, en->ex, ex->gp	.358	5	0.1222
14) en->po, gp->ex, en->pa	.220	6	0.1189
15) en->po, la->ex, ex->pa	.223	5	0.1108
16) en->po, gp->ex, po->pa	.220	6	0.1002
17) ID->po, po->en, ex->en	1.670	8	0.0931
18) en->po, la->ex, la->pa	.222	5	0.0819
19) en->po, gp->ex, la->pa	.146	7	0.0551
20) ID->ex, po->en, ex->en	1.364	8	0.0477
21) ID->ex, ID->po	1.149	11	0.0530
22) en->ex, pa->po, la->pa	.272	5	0.0438
23) en->ex, po->en			0.0336
24) en->ex, la->po, pa->la	.198	4	0.0246
25) en->ex, pa->po, en->pa	.286	6	0.0211
26) en->ex, la->po, ci->la	.198	4	0.0161
27) T->ex, T->en, T->po, T->ID, T->PD	.565	8	0.0145
28) en->ex, gp->po, pa->en	.237	4	0.0135
29) en->ex, gp->po, la->pa	.206	5	0.0133
30) en->ex, gp->po, ci->en	.237	4	0.0108
31) en->ex, pa->po, ex->pa	.286	6	0.0096
32) T->ex, T->en, T->po	2.345	7	0.0015
33) exCen, enCpo, exCpo	2.345	7	0.0015
34) gp->ex, pa->po, en->pa	.220	6	<0.001
35) ex->en, po->en			<0.001
36) gp->ex, gp->po, la->pa	.220	6	<0.001
37) gp->ex, gp->po, la->pa	.209	6	<0.001

Table 8-5: All Models Located With TETRAD

²⁵ Out of a possible 14 that hold at a significance level of .532.

8.2.4. Conclusion

It is no surprise that TETRAD's suggested sets yield Costner and Schoenberg's model immediately. The procedure by which Costner and Schoenberg obtained the model shares many features of the TETRAD algorithms and heuristics. But to justify the assumptions of the model, we either need substantive knowledge or else grounds for thinking that no other plausible model explains the data as well. We have explored only a single skeleton, but in combination with EQS, TETRAD has provided a rather full sense of the possibilities available within the boundaries of that initial hypothesis. It may be of some surprise, however, that TETRAD helps us to locate a series of plausible models that are not extensions of Costner and Schoenberg's model, that all include $en \rightarrow po$ and $en \rightarrow ex$, and that all have a much higher chi-square probability than Costner and Schoenberg's model. It may also be a surprise that the data so unambiguously choose a particular causal order among a particular pair of indicator variables.

One value of locating these alternatives is that they force substantive issues to the forefront, and they clarify which substantive issues are most urgent in theory assessment. If one wishes to defend a particular causal analysis of the data, say Costner and Schoenberg's model, then one is obliged to give good reasons for preferring that explanation to the many alternatives that seem statistically satisfactory. One might, for example, argue that the measured variables in this case are indices, and therefore cannot enter into direct causal relations with one another. (It might be argued to the contrary that causal relations among indices are acceptable surrogates for causal relations among the variables aggregated by the indices). Or one might turn to other kinds of data, or other samples, to make a case for one set of hypotheses as against another. Costner and Schoenberg's model for example, predicts that increases in per capita energy consumption will have no effect on power diversification or executive functioning. The best alternative models predict they will have such an effect.

This case illustrates that TETRAD's procedures are heuristic, not infallible, and says something about how good they are as rules of thumb. Sometimes models that are dramatic improvements when judged by statistical tests are models that slightly decrease rather than increase the P_i value. That is not surprising, given the rather arbitrary character of the measure. Sometimes models with very low TTR values do poorly when compared statistically with models having slightly higher TTR values. That is because models can imply constraints other than tetrad equations, and the TETRAD program does not, in its present version, take account of such constraints. Typically, however, TETRAD's indicators are strongly correlated with statistical indicators.

The procedure we have described may seem elaborate. In fact, it takes much longer to describe than to carry out. The entire TETRAD analysis took about half a day. The EQS runs took somewhat longer. A more thorough search would repeat the analysis on alternative skeletons. In particular, we could consider skeletons in which the measured variables are clustered differently. Costner and Schoenberg's model, with the PD \rightarrow en addition, contains eight distinct skeletons. All of them could be thoroughly analyzed with TETRAD in the course of a weekend, with more time required for statistical analysis of the most promising, distinct elaborations. In view of the time, effort and cost that goes into data collection, a few days spent analyzing alternative models does not seem excessive.

8.3. MEASURING THE AUTHORITARIAN PERSONALITY

8.3.1. Introduction

M. Kohn's *Class and Conformity* [58] describes several large studies that investigate the relationships among social class, attitudes, and personality structure. In one section of his book Kohn explores the relations between social class and a variety of latent factors, one of which is an "Authoritarian-Conservatism" personality trait. We are concerned here only with the measurement model for this latent factor.

Items on interview schedules used in social psychology are often interpreted as indicators of a common attitude or character trait that is not directly measured. Latent factors of this sort may be extracted by factor analysis, or the indicators may be chosen at the outset to reflect a common trait. After the correlations among the item responses due to this common factor are extracted, the item residuals are often still correlated, and these correlations are usually taken to indicate that the item responses have other common determinants besides the latent factor they are designed, or thought, to measure. The correlated residuals are assessed statistically by treating them as the result of correlated errors in item responses, without any further interpretation.

Thus Kohn (1969) extracted an "authoritarianism-conservatism" factor from his well-known National Survey data. Miller et al. [78] extracted an analogous factor from a parallel study of Polish men. Miller et al. isolate five "core" items, answers to interview questions, that are indicators of authoritarian personality in both national groups. Correlation data for these five indicators are available in an article by Schoenberg and Richtand [87] that demonstrates a quick maximum likelihood estimation program. The initial measurement model is pictured in Figure 8-12. Here q_1 - q_5 stand for the indicators and e_1 - e_5 are their disturbance terms.

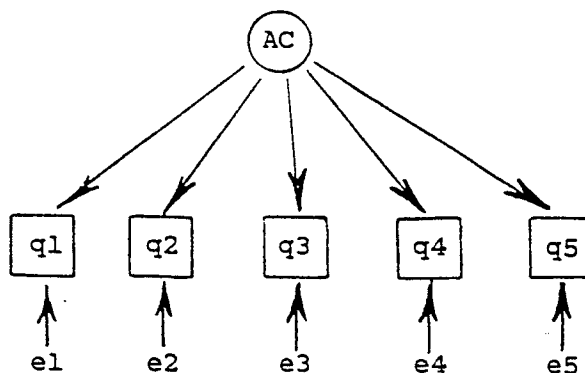


Figure 8-12: Initial AC Measurement Model

In the studies described by Kohn, approximately 3,000 subjects were presented with a number of statements. The subjects were to indicate how strongly they agreed or disagreed with these statements. The questions were scored with a five category scale from "strongly disagree" to "strongly agree." The five questions Schoenberg and Richtand list are:

1. The most important thing to teach children is absolute obedience to their parents.
2. In this complicated world, the only way to know what to do is to rely on leaders and experts.
3. No decent man can respect a woman who had sexual relations before marriage.
4. Any good leader should be strict with people under him in order to gain their respect.
5. It's wrong to do things differently from the way our forefathers did.

Miller et al. note that there are correlated residuals, and while suggesting various alternative accounts of their source, treat them statistically as correlated errors. Schoenberg and Richtand do the same. We suggest that such correlated residuals may indicate an "anchoring" effect among answers to the interview questions, analogous to that described by Kahneman and Tversky [54], and by Campbell, et al., [14]

Anchoring

An anchoring effect occurs when the subject uses the results of some mental process to serve as a starting point for another, moving from the anchor instead of starting from scratch. Kahneman and Tversky proposed such an effect to explain systematic performance errors in an experiment involving sequential multiplication tasks performed under strong time constraints. They found that subjects estimated the final product within a predictable interval of the product of the first two or three numbers. Extending these results, they found anchoring effects in a wide range of mental tasks.

A similar effect may operate in questionnaires and mental tests. It seems plausible that anchoring takes place in answering the kind of questions Kohn's study employs. Whether the subject wants to appear consistent, or answers one question with a perspective or mood already created by a previous answer, the answer to an earlier question might well affect the answer to a later one. On this assumption, we consider the hypothesis that a correlated residual between answers to interview questions is due to a *causal relation between the indicator variables*. That is, if two indicators have residual correlation not explained by the common factor already postulated, the excess correlation is due to a causal relation among the indicators, not to an unexplained correlation among their error terms.

Miller's Respecification

To interpret the residuals for items in the Kohn study and in the parallel Polish study, Miller et al. use a method that has become common in the study of measurement models. They find a maximum likelihood estimate of the factor loadings and locate the parameter (fixed at zero in the model) such that the partial derivative of the maximum likelihood fitting function is the largest for that parameter. They consider only parameters representing error correlations. They free the chosen parameter, reestimate the modified model, and find its chi square value. If the difference of chi square values between the original model and the revision is significant, the added correlation is retained. The procedure is repeated until the suggested correlations are no longer statistically significant. In this way Miller, et al. obtain the following model for the five core indicators.

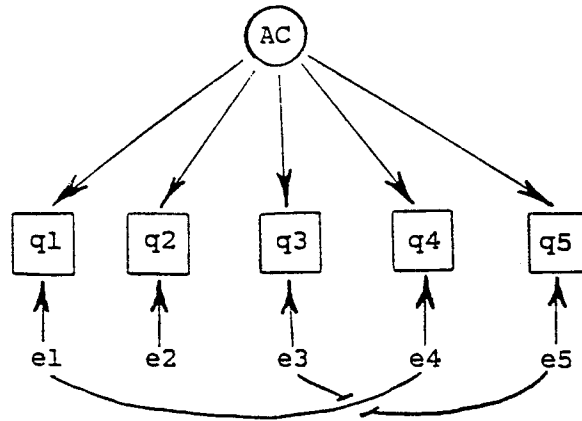


Figure 8-13: Miller's Revised Model

Schoenberg and Richtand make an identical revision, and when the factor loading of the first indicator is fixed at unity they find the probability of the chi square statistic to be 0.57.

8.3.2. The TETRAD Analysis

To locate alternative elaborations of the initial model that include causes between indicator variables, we employ TETRAD. The initial causal model is shown in Fig. 8-14.

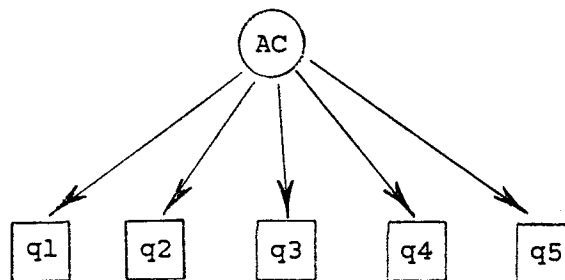


Figure 8-14: AC Skeleton

Using the Suggested Sets

We ask TETRAD for 3 units of sets of suggested sets, using the datafile anchor.dat, which is included on the TETRAD disk. The results are as follows.

The name of this output file is: anchor.ol

The graph analyzed in this example is:

AC->q1 AC->q2 AC->q3 AC->q4 AC->q5

The significance level is: 0.0500

The sample size is: 3101

Sets of suggested treks at significance level = 0.0418

{q3-q5 }

Sets of suggested treks at significance level = 0.5163

{q1-q3 q2-q5 q3-q5 }

Sets of suggested treks at significance level = 0.8099

{q1-q4 q2-q5 q3-q4 q3-q5 }
 {q1-q3 q2-q5 q3-q4 q3-q5 }
 {q1-q4 q2-q3 q3-q4 q3-q5 }
 {q1-q3 q2-q3 q2-q5 q3-q4 }
 {q1-q4 q2-q3 q2-q5 q3-q4 }
 {q2-q3 q2-q5 q3-q4 q3-q5 }
 {q1-q3 q2-q3 q2-q5 q3-q5 }
 {q1-q3 q1-q4 q2-q3 q3-q5 }
 {q1-q3 q1-q4 q2-q5 q3-q5 }
 {q1-q3 q1-q4 q3-q4 q3-q5 }
 {q1-q3 q1-q4 q2-q3 q2-q5 }
 {q1-q3 q1-q4 q2-q3 q3-q4 }

The Total Tetrad Residual, or TTR, is: 0.22699

I-H, the total number of equations implied by the model
 that do not hold in the data, is: 8

H-I, the total number of equations that hold in the data but
 are not implied by the model, is: 0

The number of equations that hold in the data is 7

The number of equations explained by the model is 7

Figure 8-15: TETRAD's Results on the AC Skeleton

Notice that TETRAD's first suggestion is the introduction of a trek between $q_3 - q_5$, two of the variables that Miller, et al. correlate. The addition of this trek does not prevent the implication of any equation that is significant at the .05 level, thus it entails virtually no loss in explanatory power. No other pairs are included in a set until the significance level is above .5, so we will use .5 as a significance level in asking for the Rtrr chart below. First, however, we pursue the suggestions TETRAD gives.

Prima facie, the second suggested set

$$\{q_1-q_3 \quad q_2-q_5 \quad q_3-q_5 \}$$

has a number of possible orderings. In accord with our theoretical considerations about anchoring, however, we only consider direct causal relationships between indicators, and this constraint makes it impossible to order the three pairs that occur in this set without creating a further trek in addition to those suggested.

For example, if we order edges between q_2-q_5 and q_3-q_5 as in Figure 8-16, then a new trek between q_2-q_3 is produced.

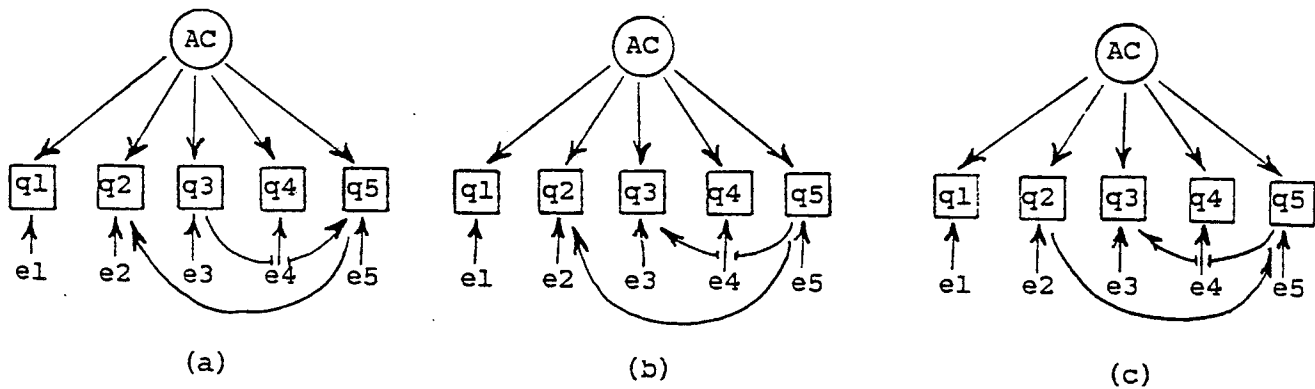


Figure 8-16: Edges That Link To Produce A Trek Between $q_2 - q_3$

If instead we add edges as in 8-17, then no new trek is created between q_2 and q_3 .

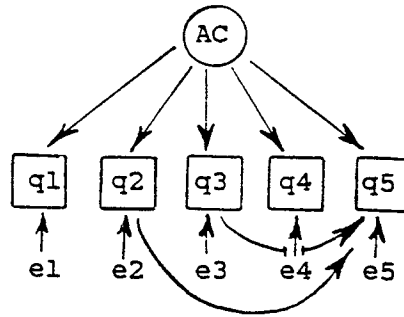


Figure 8-17:

But notice that in Figure 8-17 (and in Figure 8-16 (c)), an additional edge connecting $q1-q3$ will also create a trek between $q1$ and $q5$.

The question is: do we want to introduce a trek between $q2-q3$, between $q1-q5$, both, or neither? Looking at the next level of suggested sets, we notice that there is no set with $q1-q5$, but there are sets with $q2-q3$ in them. Using TETRAD's Compare command we find that the introduction of a new trek between $q1-q5$, $q4-q5$, $q1-q2$, or $q2-q4$ will prevent the implication of an equation that holds at a significance level greater than .91. Combinations of edges that produce these treks are to be avoided. There is a set with $q2-q3$ in the third level of suggested sets that is also a superset of the one suggested set in the second level, namely:

$$\{q1-q3 \quad q2-q3 \quad q2-q5 \quad q3-q5 \}$$

If we directly connect the three pairs that occur in the set from the second level to produce all and only the four treks in the above set, we are constrained to two possible orderings of the causal connections. Any other ordering would introduce a trek between $q1-q5$, or fail to produce a trek between $q2-q3$. The permissible hypotheses are these:

- 1) $q1 \rightarrow q3, \quad q2 \rightarrow q5, \quad q5 \rightarrow q3$
- 2) $q1 \rightarrow q3, \quad q5 \rightarrow q2, \quad q5 \rightarrow q3$

Either of these additions, or subsets of them, should do well on a chi-square test. They also contain information relevant to predicting the order in which the interview questions were asked. We cover that topic below.

Using the Rtrr Chart

Using a significance level of .5, we ask TETRAD for statistical information about each TETRAD equation and an Rtrr chart of the skeleton. The statistical information is shown in Figure 8-18 below. The Rtrr chart for the AC Skeleton is shown in Figure 8-19 below.

The graph analyzed in this example is:
AC->q1 AC->q2 AC->q3 AC->q4 AC->q5

The significance level is: 0.5000

The sample size is: 3101

Tetrad Equation	Residual	Impld.	Held	P(diff.)
q1 q2, q3 q4 = q1 q3, q2 q4	0.0068	y		0.3185
q1 q2, q4 q3 = q1 q4, q2 q3	0.0019	y	y	0.8099
q1 q3, q4 q2 = q1 q4, q3 q2	0.0087	y		0.2241
q1 q2, q3 q5 = q1 q3, q2 q5	0.0255	y		0.0000
q1 q2, q5 q3 = q1 q5, q2 q3	0.0295	y		0.0000
q1 q3, q5 q2 = q1 q5, q3 q2	0.0039	y	y	0.5163
q1 q2, q4 q5 = q1 q4, q2 q5	0.0160	y		0.0464
q1 q2, q5 q4 = q1 q5, q2 q4	0.0007	y	y	0.9193
q1 q4, q5 q2 = q1 q5, q4 q2	0.0167	y		0.0285
q1 q3, q4 q5 = q1 q4, q3 q5	0.0382	y		0.0000
q1 q3, q5 q4 = q1 q5, q3 q4	0.0051	y		0.3811
q1 q4, q5 q3 = q1 q5, q4 q3	0.0331	y		0.0000
q2 q3, q4 q5 = q2 q4, q3 q5	0.0204	y		0.0016
q2 q3, q5 q4 = q2 q5, q3 q4	0.0070	y		0.2063
q2 q4, q5 q3 = q2 q5, q4 q3	0.0134	y		0.0418

Figure 8-18: Statistical Information For the AC Skeleton

Edge	Rttr	D(I-H)	I(H-I)
q1-> q2	0.080	4	2
q2-> q1	0.080	4	2
q1 C q2	0.080	4	2
q1-> q3	0.088	5	1
q3-> q1	0.088	5	1
q1 C q3	0.088	5	1
q1-> q4	0.115	5	1
q4-> q1	0.115	5	1
q1 C q4	0.115	5	1
q1-> q5	0.089	4	2
q5-> q1	0.089	4	2
q1 C q5	0.089	4	2
q2-> q3	0.071	4	2
q3-> q2	0.071	4	2
q2 C q3	0.071	4	2
q2-> q4	0.067	5	1
q4-> q2	0.067	5	1
q2 C q4	0.067	5	1
q2-> q5	0.083	5	1
q5-> q2	0.083	5	1
q2 C q5	0.083	5	1
q3-> q4	0.067	5	1
q4-> q3	0.067	5	1
q3 C q4	0.067	5	1
q3-> q5	0.160	6	0
q5-> q3	0.160	6	0
q3 C q5	0.160	6	0
q4-> q5	0.087	5	1
q5-> q4	0.087	5	1
q4 C q5	0.087	5	1

The Total Tetrad Residual, or TTR, is: 0.22699

I-H, the total number of equations implied by the model that do not hold in the data, is: 12

H-I, the total number of equations that hold in the data but are not implied by the model, is: 0

The number of equations that hold in the data is 3

The number of equations explained by the model is 3

Figure 8-19: Rttr Chart For AC Skeleton at .5

The Rtrr chart shows that a directed edge or correlated error for q3-q5 is clearly the best candidate for addition. These treks reduce the TTR the most and cost the least in explanatory power. Notice that q1-q4, Miller's other addition, seems to be the second best candidate. We don't know how q3-q5 and q1-q4 will interact, however, and so we add only q3->q5 and then q5->q3, asking for Rtrr charts in both cases.

The graph analyzed in this example is:

q3->q5 AC->q1 AC->q2 AC->q3 AC->q4 AC->q5

The significance level is: 0.5000

The sample size is: 3.1E+03

Edge	Rttr	D(I-H)	I(H-I)	Pi
q1-> q2	0.025	2	2	0.490
q2-> q1	0.025	2	2	0.490
q1 C q2	0.025	2	2	0.490
q1-> q3	0.042	4	2	0.534
q3-> q1	0.042	4	2	0.534
q1 C q3	0.025	3	1	0.488
q1-> q4	0.043	3	1	0.538
q4-> q1	0.043	3	1	0.538
q1 C q4	0.043	3	1	0.538
q1-> q5	0.026	2	2	0.493
q5-> q1	0.042	4	2	0.534
q1 C q5	0.026	2	2	0.493
q2-> q3	0.054	4	2	0.567
q3-> q2	0.054	4	2	0.567
q2 C q3	0.022	2	2	0.480
q2-> q4	0.033	3	1	0.510
q4-> q2	0.033	3	1	0.510
q2 C q4	0.033	3	1	0.510
q2-> q5	0.044	3	1	0.538
q5-> q2	0.054	4	2	0.567
q2 C q5	0.044	3	1	0.538
q3-> q4	0.038	4	2	0.522
q4-> q3	0.038	4	2	0.522
q3 C q4	0.021	3	1	0.478
q4-> q5	0.029	3	1	0.499
q5-> q4	0.038	4	2	0.522
q4 C q5	0.029	3	1	0.499

The Total Tetrad Residual, or TTR, is: 0.06700

The Pi value for the model is: 0.56348

Figure 8-20: Rttr Chart For AC Skeleton + q3->q5

The graph analyzed in this example is:
 q5->q3 AC->q1 AC->q2 AC->q3 AC->q4 AC->q5

Edge	Rttr	D(I-H)	I(H-I)	Pi
q1-> q2	0.025	2	2	0.490
q2-> q1	0.025	2	2	0.490
q1 C q2	0.025	2	2	0.490
q1-> q3	0.025	3	1	0.488
q3-> q1	0.042	4	2	0.534
q1 C q3	0.025	3	1	0.488
q1-> q4	0.043	3	1	0.538
q4-> q1	0.043	3	1	0.538
q1 C q4	0.043	3	1	0.538
q1-> q5	0.042	4	2	0.534
q5-> q1	0.042	4	2	0.534
q1 C q5	0.026	2	2	0.493
q2-> q3	0.022	2	2	0.480
q3-> q2	0.054	4	2	0.567
q2 C q3	0.022	2	2	0.480
q2-> q4	0.033	3	1	0.510
q4-> q2	0.033	3	1	0.510
q2 C q4	0.033	3	1	0.510
q2-> q5	0.054	4	2	0.567
q5-> q2	0.054	4	2	0.567
q2 C q5	0.044	3	1	0.538
q3-> q4	0.038	4	2	0.522
q4-> q3	0.021	3	1	0.478
q3 C q4	0.021	3	1	0.478
q4-> q5	0.038	4	2	0.522
q5-> q4	0.038	4	2	0.522
q4 C q5	0.029	3	1	0.499

The Total Tetrad Residual, or TTR, is: 0.06700

The Pi value for the model is: 0.56348

Figure 8-21: Rttr Chart For AC Skeleton + q5->q3

Notice that after either q3->q5 or q5->q3 has been added, q1-q4 is no longer the addition that most reduces the TTR. We are considering only directed edges, not correlated errors, and we find that the q2-q5 edges do most to reduce the TTR, but they increase H-I by 2 whereas q1-q4 edges increase H-I by 1. Only the q2-q5 edges result in an increase in the Pi value compared with the previous models. If we did not already know

the properties of Miller's model, we would certainly include $q1-q4$ as a node in the second level. In this case, however, we will confine attention to the $q2-q5$ connections and form the following tree:

Root	Level 1			Level 2		
-----	-----			-----		
	Edge	TTR	PI			

Initial Model	($q3 \rightarrow q5$)	.067	.563	1)	($q2 \rightarrow q3$)	.013 .567
				2)	($q3 \rightarrow q2$)	.013 .567
				3)	($q5 \rightarrow q2$)	.013 .567
				4)	($q3 \rightarrow q2$)	.013 .567
	($q5 \rightarrow q3$)	.067	.563	5)	($q2 \rightarrow q5$)	.013 .567
				6)	($q5 \rightarrow q2$)	.013 .567

Figure 8-22:

Pi values for the Rtrr charts of models on level 2 indicate that searching for additional edges is not worthwhile. These nodes look identical according to TETRAD's Rtrr chart. To confirm this we use TETRAD's "Compare" option. It tells us exactly which equations previously implied will no longer be implied if an edge is added, and it tells us the residual and the probability for each of these. Adding either $q3 \rightarrow q5$ or $q5 \rightarrow q3$ to the skeleton defeats the implication of the same equations. We compare the equations defeated by node 1 of level 2, which is the skeleton + $q3 \rightarrow q5$ + $q2 \rightarrow q3$, and node 6 of level 2, which is the skeleton + $q5 \rightarrow q3$ + $q5 \rightarrow q2$.

The graph analyzed in this example is:
 $q3 \rightarrow q5$ AC- $\rightarrow q1$ AC- $\rightarrow q2$ AC- $\rightarrow q3$ AC- $\rightarrow q4$ AC- $\rightarrow q5$

The edge added was $q2 \rightarrow q3$

Tetrad Equation	Residual	Impld.	Held	P(diff.)
$q1 q2, q4 q3 = q1 q4, q2 q3$	0.0019		y	0.8099
$q1 q3, q4 q2 = q1 q4, q3 q2$	0.0087			0.2241
$q1 q3, q5 q2 = q1 q5, q3 q2$	0.0039		y	0.5163
$q1 q4, q5 q2 = q1 q5, q4 q2$	0.0167			0.0285
$q1 q2, q4 q5 = q1 q4, q2 q5$	0.0160			0.0464
$q2 q3, q5 q4 = q2 q5, q3 q4$	0.0070			0.2063

The graph analyzed in this example is:
 $q5 \rightarrow q3$ AC- $\rightarrow q1$ AC- $\rightarrow q2$ AC- $\rightarrow q3$ AC- $\rightarrow q4$ AC- $\rightarrow q5$

The edge added was $q5 \rightarrow q2$

Tetrad Equation	Residual	Impld.	Held	P(diff.)
$q1 q2, q4 q3 = q1 q4, q2 q3$	0.0019		y	0.8099
$q1 q3, q4 q2 = q1 q4, q3 q2$	0.0087			0.2241
$q1 q3, q5 q2 = q1 q5, q3 q2$	0.0039		y	0.5163
$q1 q4, q5 q2 = q1 q5, q4 q2$	0.0167			0.0285
$q1 q2, q4 q5 = q1 q4, q2 q5$	0.0160			0.0464
$q2 q3, q5 q4 = q2 q5, q3 q4$	0.0070			0.2063

Figure 8-23: Compare For Two Nodes of Level 2

The implication of the same equations is prevented by either addition. It should be emphasized that the fact that two models imply the same tetrad equations does not mean that they will produce equivalent Rtr charts. We will not pursue the search tree further.

8.3.3. Goodness of Fit Results

Running all the models we have discussed through EQS/PC, we obtain the following results.

Edge Additions To Skeleton	TTR	p(χ^2)
None (Skeleton)	.2269	<.001
Miller's Revision		
1) q3 C q5, q1 C q4	.0240	0.585
From the Suggested Sets		
1) q5→q3, q2→q5, q1→q3		0.994
1) q5→q3, q5→q2, q1→q3		0.994
From the Rtrr Chart		
1) q3→q5, q2→q3	.0130	0.098
2) q3→q5, q3→q2	.0130	0.098
3) q3→q5, q5→q2	.0130	0.666
4) q5→q3, q3→q2	.0130	0.098
5) q5→q3, q2→q5	.0130	0.666
6) q5→q3, q5→q2	.0130	0.666

Figure 8-24: Goodness Of Fit Results: All Models

8.3.4. Predicting Temporal Order

The TETRAD analysis was carried out with a pilot version of the TETRAD program before we knew the order in which the questions were asked in Kohn's survey. The results of the analysis imply predictions about that order. Earlier questions can have causal effects on later questions but not vice-versa.

We present our analysis of the evidence for a prediction about the temporal ordering among the item responses in stages of increasingly stronger predictions.

Given three questions, there are six possible orders in which the questions can be asked. An intermediate stage of the TETRAD analysis, nodes 3, 5 and 6 of level 2 of the Rtrr search, permitted only three possible causal relations:

- 1) q3 → q5 q5 → q2
- 2) q5 → q3 q5 → q2
- 3) q5 → q3 q2 → q5

These relations are consistent with only four possible temporal orderings of questions 3, 2 and 5. They require that question 5 either precede both of questions 2 and 3 or else occur between them; that is to say the possible temporal orderings are:

$$\begin{array}{l} 3 - 5 - 2 \\ 2 - 5 - 3 \\ 5 - 2 - 3 \\ 5 - 3 - 2 \end{array}$$

We are quite confident of this conclusion, which has an a priori chance of 2/3.

If we now consider the addition of a 1-3 trek, as in the second level of suggested sets, with the restriction that a trek must not be introduced between 1 and 5, the first of the preceding causal orderings is eliminated. Thus our heuristic procedures led us to the conclusion that the only admissible possible causal arrangements of q5, q3 and q2 is:

$$\begin{array}{l} 2) \quad q5 \rightarrow q3 \quad q5 \rightarrow q2 \\ 3) \quad q5 \rightarrow q3 \quad q2 \rightarrow q5 \end{array}$$

These relations are consistent with only three possible temporal orderings of questions 3, 2 and 5, namely:

$$\begin{array}{l} 5 - 3 - 2 \\ 5 - 2 - 3 \\ 2 - 5 - 3 \end{array}$$

In view of the consilience, we are fairly confident that the actual order in which the questions were asked was one of these three. The hypothesis has an a priori chance of 1/2.

The only permissible way to add the 1-3 trek given our other hypotheses is as $q1 \rightarrow q3$. Thus we obtain the two orderings of $q1, q3, q5, q2$ found from the Suggested Sets procedure. There are still only two possible arrangements of the directed edges permitted

by the TETRAD analysis, but they involve four variables. For four questions there are twenty four possible orders in which the questions might have been asked. The TETRAD suggestion is consistent with only eight of them:

1 - 5 - 3 - 2
 5 - 1 - 3 - 2
 1 - 5 - 2 - 3
 5 - 1 - 2 - 3
 5 - 2 - 1 - 3
 1 - 2 - 5 - 3
 2 - 1 - 5 - 3
 2 - 5 - 1 - 3

The hypothesis that the actual ordering of the questions was one of these has a prior probability of 1/3, and we have some confidence in the hypothesis.

These predictions were made before we knew the actual ordering of the questions. In fact, it was because we realized that the TETRAD program and our heuristics had yielded conclusions about causal order contrary to the numbering of the questions in Schoenberg and Richtand's report²⁶ that we understood that the program was in fact making claims about the order in which the questions were asked.

The actual order of the questions on the National Survey is

2 - 1 - 4 - 5 - 3

which is one of the possible orderings of the five questions allowed by our strongest prediction.

8.3.5. Conclusion

The data support an interpretation of the correlated residuals as direct causal effects between indicator variables. Not only do our best models do extremely well on a standard statistical test, they also yield a correct prediction about the order of the interview questions.

²⁶ Nothing Schoenberg and Richtand say suggests that the numbering used in their report was the same as the ordering for the initial study reported by Kohn.

8.4. ALTERNATIVES TO REGRESSION MODELS

8.4.1. Introduction

Regression models are among the most common linear statistical theories applied to non-experimental data. They are often applied to "demonstrate" the effect of one or more "independent" variables on a "dependent" variable. The terminology implies strong causal assumptions, and in many cases the implication is deliberate. Graphically, regression models may have one or several source variables, which may or may not be correlated, and they contain directed edges from each source variable to the "dependent" variable:

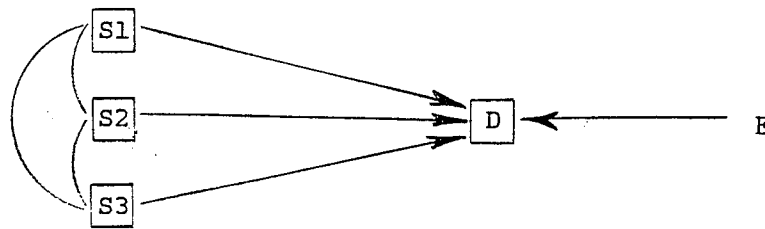


Figure 8-25: Graph of a Regression Model

Although presuppositions of regression models, such as the linearity of the dependencies, or normality, can be tested, the models imply no constraints on the correlations, and so cannot be tested more directly. Estimates of the regression coefficients are usually accompanied by t tests in which the null hypothesis is that some regression coefficient is zero. These tests are, however, no more than tests of the existence of a particular causal edge *on the assumption* that the remainder of the regression model is correct. Unlike chi square tests of overidentified models, t tests are not tests of the regression model as a whole. T tests represent a kind of "bootstrap" test, not in the statistician's sense but in a more general sense (see Glymour, [34]). Part of the theory is assumed in testing one hypothesis within the theory; in turn, the hypothesis tested is assumed in testing other hypotheses in the part of the theory that were presupposed in the first test. The result is a kind of circularity which is not vicious, but which is often not very powerful either. The bootstrap style t tests of regression models do not succeed in eliminating the possibility that there are other theories that will better explain the data.

Because regression models are not testable as a whole, one cannot make a case for such a model based on the fact that it has passed a statistical test that is powerful against other

alternative causal hypotheses. And for the same reason, because regression models do not imply constraints on the correlation matrix, one cannot make a case for a regression model on the grounds that it explains such constraints. It is therefore very difficult to argue that a regression model provides the best explanation of the data, unless prior knowledge guarantees that the implicit causal assumptions are correct. When the variables are indexed by different times, as in econometric time series, that is sometimes possible. More often, however, regression models contain causal assumptions that are not established by prior knowledge, and there is no excuse for failing to consider alternative causal explanations of the data. When there are strong constraints that show up in the sample data, and there are alternative causal theories that can explain those constraints, can pass statistical tests, and are consistent with our prior knowledge, then the alternative causal models may very well be preferable to the regression model. TETRAD can be used to help find such alternatives.

8.4.2. The Dependency Theory Example

Timberlake and Williams, [105], claim that foreign investment in third-world or "peripheral" nations, causes the exclusion of various groups from the political process within a "peripheral" country. Put more simply, foreign investment promotes dictatorships and oligarchies. They also claim that "foreign investment penetration increases government repression in noncore countries" (p. 144). It is clear that such theses, if true, have important policy implications. Timberlake and Williams try to support their first claim by means of a simple regression model. Their more complicated argument for the second thesis depends on the correctness of the regression model they propose. We will concentrate on their regression model and on alternatives to it.

Timberlake and Williams develop measures of political exclusion, foreign investment penetration (in 1973), energy development, civil liberties, population, government sanctions in two years (1972 and 1977) and political protests in those same years. They correlate these measures for 72 "non-core" countries. All of the variables, save population, have substantial positive or negative correlations with one another, with absolute values ranging from .123 to .864. It should be noted that their investment data concern a period preceding the increase in petrodollars loaned to third world countries following the dramatic OPEC increases in oil prices.

A straightforward embarrassment to the theory is that political exclusion is *negatively* correlated with foreign investment penetration, and foreign investment penetration is *positively* correlated with civil liberties and negatively correlated with government sanctions. Everything appears to be just the opposite of what the theory requires. The gravamen of the Timberlake and Williams argument is that these correlations are misleading, and when other appropriate variables are controlled for, the effects are reversed.

To sustain the first hypothesis, they regress the political exclusion variable on foreign investment penetration together with energy development and civil liberties (measured on a scale whose increasing values measure decreases in civil liberties). Pictorially,

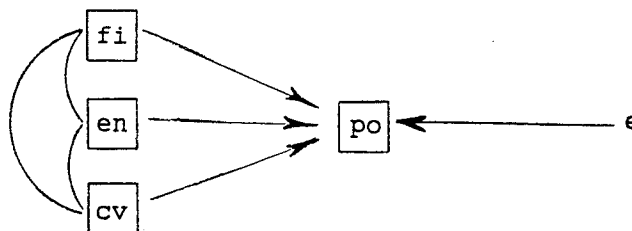


Figure 8-26: T & W's First Hypothesis

They find a statistically significant positive regression coefficient for foreign investment penetration and conclude that their first hypothesis is supported.

Timberlake and Williams thus claim to have found evidence that foreign investment in third-world nations causes governments to be unrepresentative and undemocratic. Their conclusion implies that the development of democracy and human rights would have been furthered in the early 1970s if international corporations, the world bank, private banks and other organizations based in industrial countries had not invested in third world nations. But what evidence has the regression analysis actually provided?

The analysis *assumes* that political exclusion is the effect of the absence of civil liberties, of energy development and of foreign investment. They further assume that these causes act independently, that their effects are additive, and that nothing else has an effect on any of the independent variables and on political exclusion. They give no particular reasons for these assumptions, and one might have thought otherwise. For example, one might have thought that unrepresentative government causes an absence of civil liberties, or each causes the other. *Absolutely nothing* has been done to show that the regression model and its accompanying causal assumptions are correct. But Timberlake and Williams do nothing further to establish it. There really is no argument for the regression model, or for the more particular thesis that foreign investment promotes political exclusion, other than the "bootstrap" tests provided by the t tests of linear coefficients.

8.4.3. The TETRAD Analysis

There are some puzzling features of the data, which we might expect a good theory to explain. For example, there are in the data some relations among the correlations that hold much more exactly than we expect by chance. Using TETRAD we find that the following relations hold almost exactly in the sample data:

A.

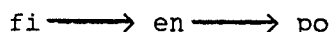
$$\rho_{po,fi} - \rho_{po,en}\rho_{en,fi} = 0$$

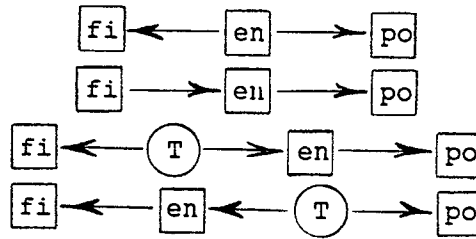
B.

$$\rho_{ev,cv} - \rho_{en,po}\rho_{po,cv} = 0$$

Actually, in the sample, the differences on the left in equations A and B are not quite zero. We can test these equations by asking the following: if there were an infinite population in which equations A and B held exactly, and we drew a random sample of size 72 (the number of "non-core" nations in the study), how likely would we be to get in that sample differences that are as large as those we find? If the answer is a very small probability, we can argue that the sample provides no evidence that the constraints A and B are other than artifacts. But in fact we find just the reverse. We find, again using TETRAD, that if the constraints A and B hold, then the probability of obtaining a difference at least as large as that found in the sample for equation A is .868, and the probability of obtaining a difference at least as large as that found in the sample for equation B is .800. These numbers help convince us that A and B are real constraints on the measured variables, and should, if possible, be explained.

These equations are interesting exactly *because they are the kind of relationship among correlations that can be explained by causal structure*. The first equation can be explained by supposing that the only effects of political exclusion on foreign investment, or of foreign investment on political exclusion, or of any third factor on both political exclusion and foreign investment, are mediated by per capita energy consumption: one variable affects another only through its effect on energy consumption. More visually, the first equation will be explained provided the causal connections between political exclusion and foreign investment are as illustrated in the next figure:





In the same way, the second equation, B, can be explained by supposing that any correlations between energy consumption and absence of civil liberties are due to the effects of political exclusion, e.g., if increases in per capita energy consumption cause an increase in civil liberties, they do so because of their direct effect on totalitarianism.

Timberlake and Williams' model does not provide any causal explanation of relations A and B, but it is easy to find assumptions that do explain these patterns, and explain them rather neatly. We will exhibit some alternative explanations pictorially.

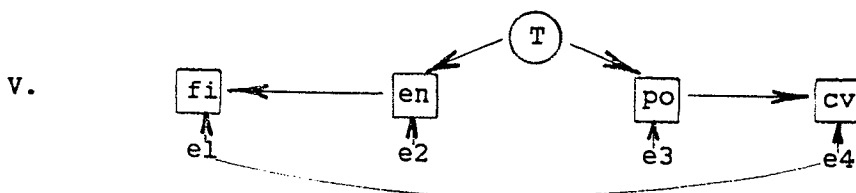
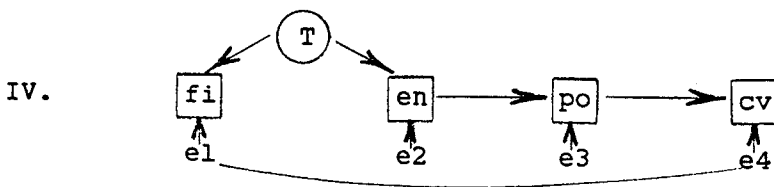
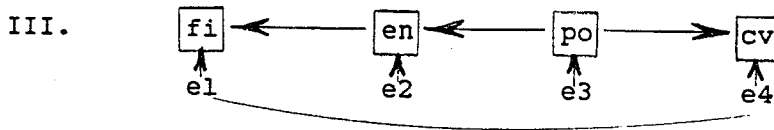
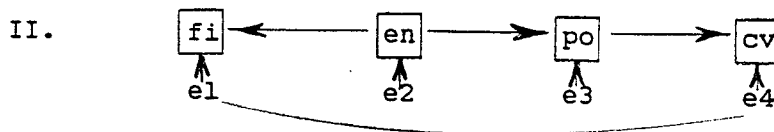
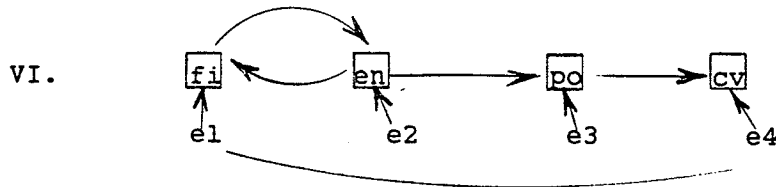


Figure 8-27: Alternatives to the Regression Model

T signifies a latent common cause. The causal hypotheses in all figures, under the assumption of linearity, imply that both A and B hold in the population, *no matter what the values of the linear coefficients may be.*

There is also a plausible non-recursive model that explains equation A but not equation B, namely



We have used the EQS program to estimate and test Model II. All linear coefficients are very significant. The coefficient giving the dependence of fi on en is positive; the coefficient giving the dependence of po on en is negative; and the coefficient giving the dependence of cv on po is positive. The p value for the chi square statistic with 2 degrees of freedom is .94.

If one accepts Model II, then the conclusion is that foreign investment in "peripheral" nations neither promotes nor inhibits the development of democracy and civil liberties, but raising the energy consumption per capita promotes both foreign investment and more representative government, and through representative government increases respect for civil liberties. We would not on this data, and given the alternatives, argue that Model II should be accepted. We do claim that it, and very likely the other alternatives suggested here, is preferable to Timberlake and Williams' regression model.

8.5. INTRODUCING LATENT VARIABLES LONGITUDINAL DATA WITH SAT SCORES

TETRAD is designed to search for elaborations of an initial model, but the program can also help to decide between alternative initial causal models. We know from an earlier chapter that tetrad equations can be derived from vanishing numerators of partial correlations, and we have suggested the following heuristic principle:

If the tetrad equations found to hold in the sample are all implied by vanishing partial correlations found to hold in the sample, then, ceteris paribus, do not introduce latent variables to account for the correlations. But, ceteris paribus, if the tetrad equations found to hold in the sample are not all implied by vanishing partial correlations found to hold in the sample, then, provided no cyclic model in the measured variables is plausible, do introduce latent variables to account for the correlations.

In the second case, the latent variables should be introduced in such a way that the model implies the tetrad equations found to hold, and does so because the structure of the model is such that measured correlations must vanish when partialled on one or another latent variable.

8.5.1. An Example

Consider data from a longitudinal study of the performances of 799 school girls on the Scholastic Aptitude Test. The same cohort of students took the test in the 5th, 7th, 9th and 11th grades (see Joreskog, [51]). The variance/covariance matrix is included on the TETRAD disc in a file named sat.dat. The matrix is:

	q5	q7	q9	q11
q5	67.951			
q7	71.01	141.578		
q9	85.966	134.748	249.748	
q11	97.153	151.068	218.757	300.669

Using the TETRAD program, we compute for each possible vanishing tetrad difference and for each possible vanishing partial correlation, the probability of the sample difference on the hypothesis that the population difference is zero. The relevant output is:

Tetrad Equation	Residual	P(diff.)
$q_5 q_7, q_9 q_1 = q_5 q_9, q_7 q_1$	0.0953	0.0000
$q_5 q_7, q_1 q_9 = q_5 q_1, q_7 q_9$	0.0917	0.0000
$q_5 q_9, q_1 q_7 = q_5 q_1, q_9 q_7$	0.0036	0.7580

Partial	Residual	P(diff.)
$q_5 q_7. q_9$	0.4806	0.0000
$q_5 q_7. q_1$	0.4542	0.0000
$q_5 q_9. q_7$	0.2931	0.0000
$q_5 q_9. q_1$	0.2655	0.0000
$q_5 q_1. q_7$	0.3177	0.0000
$q_5 q_1. q_9$	0.3379	0.0000
$q_7 q_9. q_5$	0.4595	0.0000
$q_7 q_9. q_1$	0.3208	0.0000
$q_7 q_1. q_5$	0.4742	0.0000
$q_7 q_1. q_9$	0.3819	0.0000
$q_9 q_1. q_5$	0.6347	0.0000
$q_9 q_1. q_7$	0.5763	0.0000

Figure 8-28:

Because of the time ordering of the variables, no cyclic model without latent variables is reasonable. The heuristic tells us that an adequate linear model for this data should contain a latent variable. If, for example, we instead attempt to model the data by supposing that each measurement is a direct cause of the succeeding measurement, e.g.,

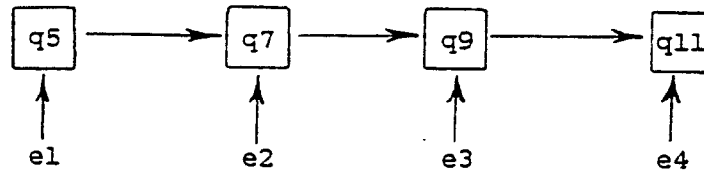


Figure 8-29: Path Model

then the model implies the third tetrad equation, and no other vanishing tetrad differences, but it also implies several vanishing partial correlations, for example, that q_5 and q_{11} vanish when partialled on q_9 . If the model is modified to avoid these incorrect constraints, for

example by correlating the error terms or by introducing further direct effects between earlier and later measurements, then the third tetrad equation is no longer implied.

If we attribute the correlations to the action of a latent variable or variables, then no vanishing partial correlations will be implied. Unless the latent structure is chosen carefully, however, the wrong tetrad constraints will be implied. For example, if the data are explained by postulating a single latent variable ("test taking ability," or whatever) and permitting it to have different linear effects on the several administrations of the SAT, then we obtain the model:

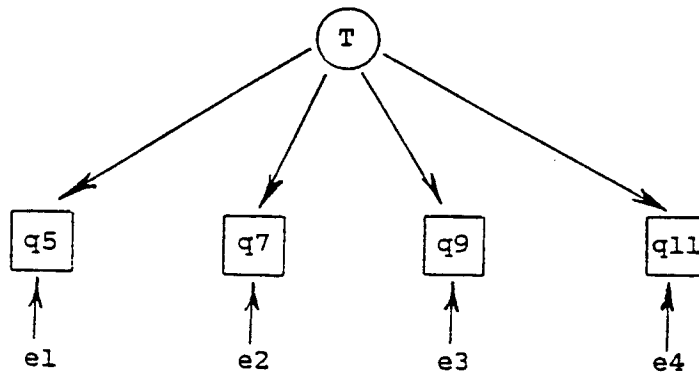


Figure 8-30: Factor Model

This model implies all three tetrad equations for the four measured variables, and judged from the sample, two of these implications are incorrect.

The simplest way to form a model that implies the single tetrad equation is to introduce latent variables that imply that the correlations in the tetrad vanish when partialled on some latent variable. There should be a latent variable such that every trek between q_5 and q_9 , between q_7 and q_{11} , between q_5 and q_{11} and between q_7 and q_9 passes through that variable. But, in the simplest case, there must be treks between q_5 and q_7 , and between q_9 and q_{11} , that do not pass through any latent variable common to all treks among the other pairs, because tetrad equations involving these latter variable pairs do not hold in the data, and should not be implied by the model.

An adequate model is obtained if we treat the latent variable as itself lagged:

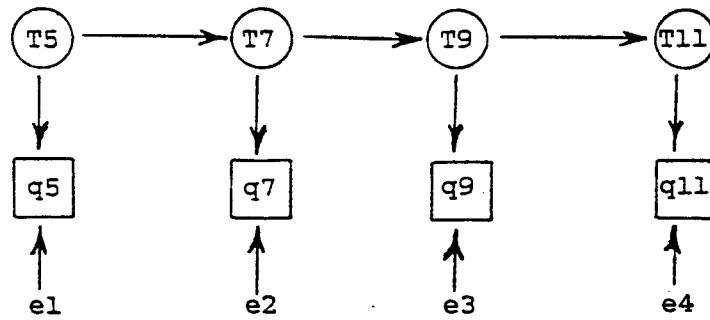


Figure 8-31: Single, Self Lagged Latent Variable Model

This model implies the tetrad equation found to hold in the sample, and only that tetrad equation. It does not imply any vanishing partial correlations among the measured variables. The model is not identifiable, but it is extremely plausible.

8.6. THE CAUSAL RELATIONS BETWEEN CONSTITUENT ATTITUDES AND CONGRESSIONAL ROLL CALL VOTING

8.6.1. Introduction

In the mid 1960s a series of articles appeared in the *American Political Science Review*, [77, 16, 25], all concerning data about constituent attitudes and congressional voting behavior. The sequence of articles is especially interesting both because of substantive issues and because of the very different methodological views of the several authors. Miller and Stokes, who originally framed the substantive question discussed in all three articles, attempt to distinguish between two possible causal paths by which constituents influence their representatives in Congress. The principle methodological problem addressed by all the papers is that of finding the best causal explanation for the correlational data, and more specifically, the problem is to distinguish a number of possible causal models that might explain the same data.

Miller and Stokes argue for a particular causal model on the basis of plausibility and regression results, and then estimate the coefficients in submodels to determine the relative strengths of the different causal pathways. Cnudde and McCrone self consciously employ a technique due to Simon and to Blalock, although they do not clearly characterize the technique.²⁷ The technique amounts to testing models by means of the overidentifying constraints they imply, and eliminating those that imply constraints found not to hold in the data. Most of the constraints considered are vanishing partial correlations, but the authors apply no statistical tests. Using a body of substantive assumptions about the existence and direction of causal relationships among a subset of the variables they model, Cnudde and McCrone distinguish among possible causal structures for the remaining variables by showing that some structures imply constraints that fail to hold in the data. Forbes and Tufte rightly demand a justification of the assumptions that limit the number of overall causal structures examined. They show that when Cnudde and McCrone's assumptions are violated there are alternative models that imply no false constraints but which lead to quite different substantive conclusions. They fail to realize the role of explanation in theory assessment, however, and they consider all models that imply no false constraint to be equivalent.

Using the TETRAD program, we show how the Simon-Blalock procedure can be quickly, conveniently, and more exactly applied, we show that latent variable models can be eliminated, and most importantly we show how models considered equivalent by Forbes and

²⁷For Simon's work, see [90, 91]. For Blalock's, see [7].

Tufte provide dramatically different levels of explanatory power.

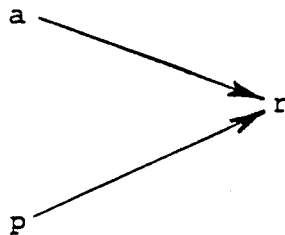
8.6.2. Miller and Stokes' Article

In 1963, Warren Miller and Donald Stokes noted a high correlation between constituent attitudes and congressional roll call voting behavior in at least two broad dimensions of policy, domestic intervention and civil rights. In later articles only data on civil rights is discussed, so we focus on that dimension. Miller and Stokes framed two possible mechanisms by which a correlation between constituent's attitudes and congressional voting could be explained. Voters might recruit candidates who share their attitudes, and those candidates who are elected then vote according to their own attitudes. Alternatively, incumbent politicians who wish to stay in office vote in accord with what they perceive to be their constituents attitudes. Their discussion and those that followed included four variables. We list them and include our scheme of abbreviation, which we take from Cnudde and McCrone:

Attitudes of the Constituents in a Representative's District	=	d
Representative's Attitudes	=	a
Representative's Perception of Constituent's Attitudes	=	p
Roll Call Vote	=	r

Figure 8-32: Variables For The Roll Call Studies

Miller and Stokes use regression studies to argue that p is a cause of r and a is a cause of r , as pictured below.



They then specify the causal model we picture in Figure 8-33 below. Their strategy is to determine the relative strengths of the causal paths from d to r . Following Sewall Wright, they attempt to analyze what proportion of r 's variance can be explained by the alternative causal pathways that correspond to the rival hypotheses they entertain.

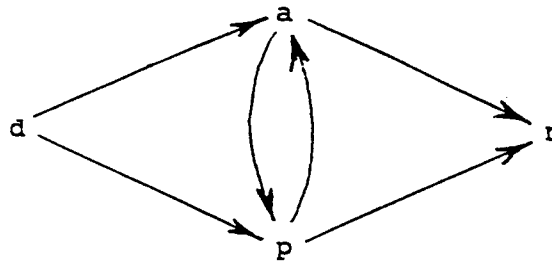


Figure 8-33: Miller and Stokes - 1963

They cannot estimate coefficients because the model they specify is underidentified. Their strategy is therefore to consider two sub-models that are identified, namely models I and II in Figure 8-34 below, and establish bounds on the values of certain coefficients.

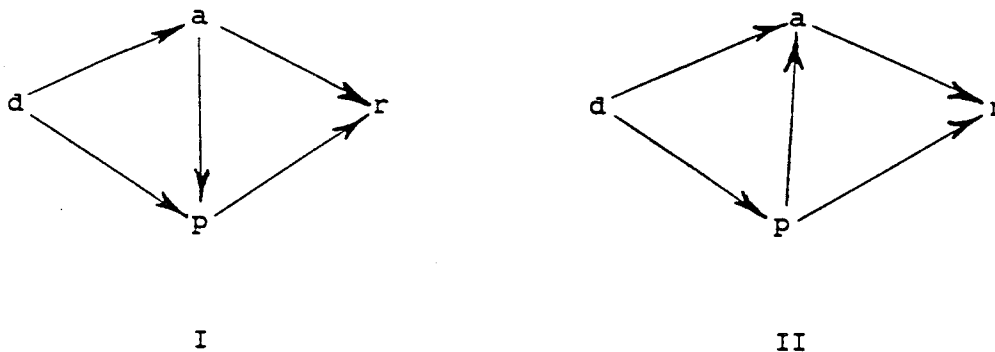


Figure 8-34: Two Sub Models

They argue that in the least favorable of these models, the sum of all the causal paths traveling from d to r that go through p explains more of the variance in r than does the sum of all the paths from d to r that go through a . Their conclusions turn on the causal model they embed the two submodels within, and this model is untestable as a whole.

8.6.3. Cnudde and McCrone

The next article in the series, by Cnudde and McCrone, takes the analysis several steps further. They summarize their strategy:

...the Simon-Blalock causal model analysis is used to resolve the two questions: are all the possible causal paths (in the Miller and Stokes model) operative? What is the direction of causation between the representative's perception and his attitude?

They quote Blalock to summarize what they take to be the essence of the Simon-Blalock technique:

to make causal *inferences* concerning the adequacy of causal models, at least in the sense that we can proceed by eliminating inadequate models that make predictions that are not consistent with the data.

Cnudde and McCrone make a host of assumptions that limit the number of models they need consider.

1. There are no cycles in the directed graph.
2. Like Miller and Stokes, they assume that a causes r and that p causes r , and that r is a cause of no variable.
3. The variable d does not directly cause the variable r .
4. The variable d is not caused by any variable.
5. The variables d , a , and p are all trek connected.²⁸

Given these assumptions, they consider models that vary only with regard to the causal ordering of d , a and p . We show three of these models, the ones Cnudde and McCrone explicitly consider, in Figure 8-35.

²⁸This simply means that there is a trek between any two.

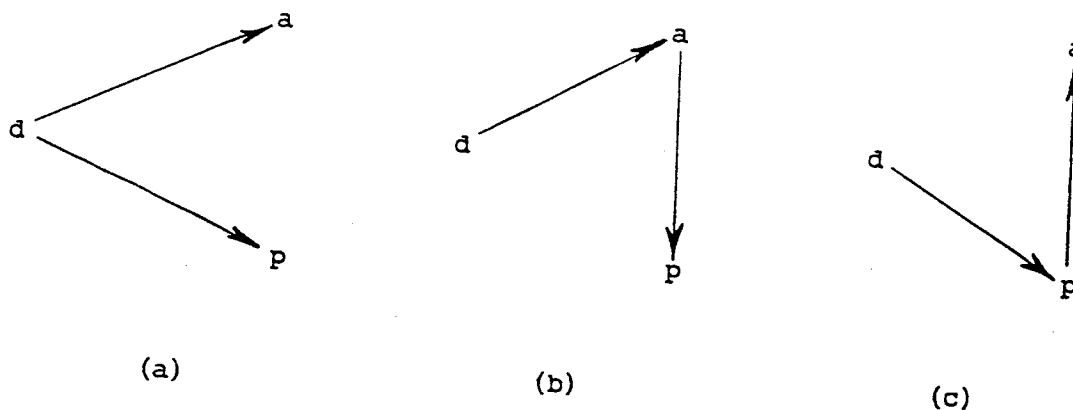


Figure 8-35: Three possible orderings among d , a , and p

The first, model (a), postulates no direct causal link between a and p . They note that this implies:

$$\rho_{da} \rho_{dp} = \rho_{ap}$$

which in turn implies (provided that neither ρ_{da} or $\rho_{dp} = 1$):

$$\rho_{ap.d} = 0$$

This prediction is false, and Cnudde and McCrone conclude that p and a have some direct causal connection. They then seek to establish which way the causal direction propagates. To do this they consider models (b) and (c). Model (b) implies that

$$\rho_{da} \rho_{ap} = \rho_{dp}$$

This prediction is also false, so this model can be ruled out. Model (c) implies that

$$\rho_{dp} \rho_{pa} = \rho_{da}$$

This equality holds closely in the data, although no statistical test is given, and the authors

conclude that this half of the model is confirmed.

They then argue that a direct connection between p and r and between a and r cannot be removed, because in either case the extension of model (c) would imply an equation that does not hold in the data. Finding that the extension of model (c), pictured in Figure 8-36, predicts no false equations, they conclude that it fits the data as well as does the Miller/Stokes Model, but they consider model (c) superior because it is more parsimonious. We would put the issue differently.

Neither of these models imply any *very false* vanishing partial correlations, but the two models nonetheless have very different implications. Miller and Stokes' model implies no false constraints because it implies no constraints at all. The Cnudde model implies a constraint that happens to hold. To imply a constraint that holds closely is to explain it. The models are not equivalent on explanatory grounds, and model (c) is preferable.

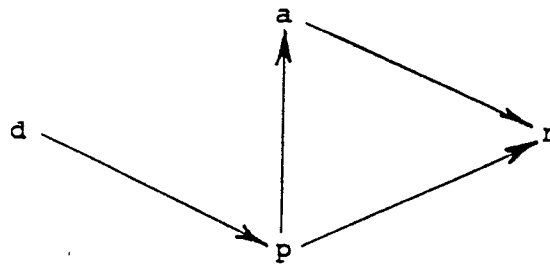


Figure 8-36: Cnudde and McCrone's Suggested Model

8.6.4. Forbes and Tufte

Hugh Forbes and Edward Tufte's paper turns the confusion concerning equivalence among models just noted into a disaster. The point of their study is principally methodological. They use the roll call voting data to illustrate the need for caution when making conclusions based on causal models. The principal problem they emphasize is the same problem we belabor. *One must consider alternative causal models.* They attempt to show that Cnudde and McCrone have neglected to consider alternative causal models that are indistinguishable by "the Simon-Blalock technique" and yet lead to very different substantive conclusions. They are entirely right that alternative explanations must be considered, but they think there is no reason to have any preference between models that imply *no* constraints (and

therefore no false ones), and models that imply no false constraints but that do imply *empirically correct* constraints on the data. We have argued at length in earlier chapters that this distinction is fundamental to the comparison of alternative explanations, and is one of the keys to the success of the natural sciences.

Forbes and Tufte offer the following three causal models, all of which they consider empirically indistinguishable.

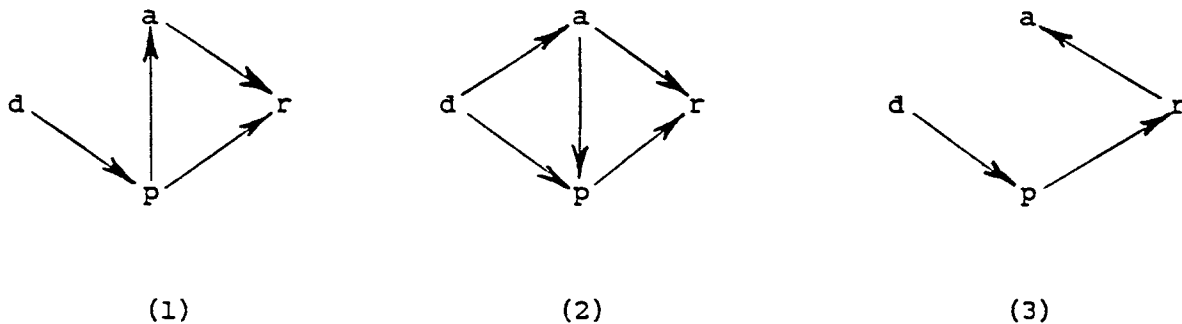


Figure 8-37: Three models from Forbes and Tufte

They say the following about these models.

Cnudde and McCrone have shown that model (1) fits the data; model (2) also fits: the addition of the link between district opinion and Congressmen's attitudes generates no new prediction equations that distinguish model (2) from (1). And model (3) also fits... [25], p. 1260.

They go on to insist that these models are distinguishable only by reliance on vague hunches and the like.

How do we choose between these three different models, each of which suggests a different conclusion about the relationship between attitudes and perceptions? It should be apparent that the choice between them must rest (in the absence of additional data) on the investigator's hunch about what causal mechanisms are likely to exist in the real world. The only other imaginable basis for choice is the criterion of parsimony... [25], p.1260.

Some models fit the data because they imply constraints that hold in the data, and explain why those constraints hold. Other models fit the data because they could fit *any possible*

data: they have a lot of parameters to gerrymander. We imagine that this difference is a very good basis for choice.

8.6.5. The TETRAD Analysis

We ran TETRAD on all three of the models in Figure 8-37, asking for statistical information about partial correlation equations and tetrad equations. We show the output for each in successive Figures below.

The graph analyzed in this example is:

d → p p → a p → r a → r

The sample size is: 116

Partial	Residual	Impld.	Held	P(diff.)
d p . a	0.6291			0.0000
d p . r	0.4718			0.0000
d a . p	0.0454	y	y	0.6253
d a . r	0.0570		y	0.5470
d r . p	0.1086	y	y	0.2485
d r . a	0.4825			0.0000
p a . d	0.4708			0.0000
p a . r	0.1261		y	0.1776
p r . d	0.6701			0.0000
p r . a	0.6772			0.0000
a r . d	0.6030			0.0000
a r . p	0.4409			0.0000

Tetrad Equation	Residual	Impld.	Held	P(diff.)
d p , a r = d a , p r	0.1222			0.0136
d p , r a = d r , p a	0.1148			0.0040
d a , r p = d r , a p	0.0075	y	y	0.8388

Figure 8-38: TETRAD Output on Model (1)

The graph analyzed in this example is:

d -> p d -> a p -> r a -> p a -> r

The sample size is: 116

Partial	Residual	Impld.	Held	P(diff.)
d p . a	0.6291			0.0000
d p . r	0.4718			0.0000
d a . p	0.0454		y	0.6253
d a . r	0.0570		y	0.5470
d r . p	0.1086		y	0.2485
d r . a	0.4825			0.0000
p a . d	0.4708			0.0000
p a . r	0.1261		y	0.1776
p r . d	0.6701			0.0000
p r . a	0.6772			0.0000
a r . d	0.6030			0.0000
a r . p	0.4409			0.0000
Tetrad Equation				
d p , a r = d a , p r	0.1222			0.0136
d p , r a = d r , p a	0.1148			0.0040
d a , r p = d r , a p	0.0075		y	0.8388

Figure 8-39: TETRAD Output on Model (2)

The graph analyzed in this example is:

d → p p → r r → a

The sample size is: 116

Partial	Residual	Impld.	Held	P(diff.)
d p . a	0.6291			0.0000
d p . r	0.4718			0.0000
d a . p	0.0454	Y	Y	0.6253
d a . r	0.0570	Y	Y	0.5470
d r . p	0.1086	Y	Y	0.2485
d r . a	0.4825			0.0000
p a . d	0.4708			0.0000
p a . r	0.1261	Y	Y	0.1776
p r . d	0.6701			0.0000
p r . a	0.6772			0.0000
a r . d	0.6030			0.0000
a r . p	0.4409			0.0000

Tetrad Equation	Residual	Impld.	Held	P(diff.)
d p , a r = d a , p r	0.1222			0.0136
d p , r a = d r , p a	0.1148			0.0040
d a , r p = d r , a p	0.0075	Y	Y	0.8388

Figure 8-40: TETRAD Output for Model (3)

One can see at a glance that none of the three models implies a false equation, but all three models are easily distinguished in what they do imply. Model (1) explains two of the four vanishing partial correlations that hold at a significance level of .177 or higher and it explains the tetrad equation that holds at a significance level above .83. Model (2), which Forbes and Tufte cannot imagine a way to distinguish from model (1), implies *no* equations. Model (3) does even better than model (1), for it implies all and only those equations that hold at .17 or higher.

If the only property of a model that matters is that it not imply a false equation, then these models are indeed indistinguishable, but then we must include a host of other models in this class. The complete graph among four variables, for example. A causal model should explain the correlations (Thurstone's Principle), but it should also explain the constraints those correlations satisfy (Spearman's Principle). Only two of the three models Forbes and Tufte consider, and only a handful of the possible models, have this property. By paying attention to a model's explanatory power, we give ourselves an immense

advantage in distinguishing among otherwise equivalent models.

Latent Variable Models

None of the models considered in this series of articles entertains the possibility that the correlations are produced by unmeasured variables. With variables as deeply imbedded in social context as are those in this data set, the action of unmeasured factors cannot be dismissed without consideration. Forbes and Tufte's emphasis on considering alternatives is appropriate, but too narrowly applied. Besides the alternatives they consider, there are any number of latent variable models that might account for the data. But for this data, latent variable models can be dismissed for the same reasons that Forbes and Tufte's alternatives can be dismissed: they do not explain the constraints the data satisfy.

Recall our principles for introducing latent variables, and their justification. When vanishing partial correlations hold that imply the tetrad equations that hold, do not introduce latent variables without good substantive reasons to do so. The latent variable models will not explain the vanishing partial correlations, and models without latent variables that imply the vanishing partial correlations will also imply the vanishing tetrad difference.

TETRAD shows us that in the roll call voting data, one tetrad equation holds, namely:

$$d_{a,r}p = d_{r,a}p$$

Its p value is .8388. Two partials, e.g.:

$$d_{a.p} \quad \text{and} \quad d_{r.p}$$

hold with p values of .62 and .24 respectively, and are also explained by models (1) and (3). These vanishing partials imply the single tetrad equation that holds, and any models that explains these vanishing partial correlations will also explain the vanishing tetrad difference. The introduction of a latent variable that still allows the graph to imply the above tetrad equation would not allow the graph to explain these two partials. If we were to introduce a latent variable as a common cause of either d-a, d-r, or p-a, then we would prevent the explanation of the vanishing partial correlation equations involving these pairs.

8.6.6. Goodness of Fit Results

Before exhibiting the results of a chi-square test on all the models we have considered so far, we would like consider one more alternative.

Model (3) in Figure 8-37 is difficult to believe. It asserts that a representative's roll call vote is a causal factor in the formation of his or her attitudes, but that a representative's attitudes have no influence at all on his or her roll call vote, as though a representative discovers what he or she thinks only by observing his or her own voting behavior. It seems rather more plausible to simply add a cause from a to r, allowing the relationship between a and r to be cyclic. TETRAD shows that the resulting model, pictured in Figure 8-41, explains the same vanishing partial correlations as does model (1) in Figure 8-37, although it is less simple. The two models are not, however, statistically equivalent.

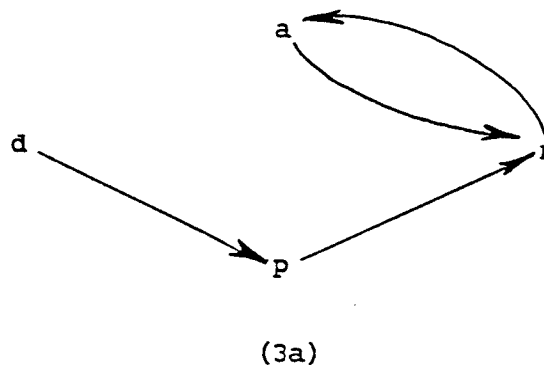


Figure 8-41: Final Candidate

We ran the model above and those shown in Figure 8-37 through EQS. The results are given below.

Model	DF	X ²	p(X ²)
(1)	2	1.365	0.505
(2)	1	1.128	0.288
(3)	3	3.207	0.361
(3a)	2	1.365	0.505

Figure 8-42: Goodness Of Fit Results

8.6.7. Conclusion

All of these models perform acceptably under the chi square test, although model (2), which explains no equations, does markedly worse than the others.

Chudde and McCrone's procedure, like Blalock's, shares the methodological sensibilities behind the TETRAD program. The program makes it possible to do this kind of analysis much more systematically, however, and to distinguish among a much broader class of alternative models. It also supplies statistical tests in place of informal judgements about vanishing partial correlations and vanishing tetrad differences.

8.7. THE EFFECTS OF SUMMER HEAD START

8.7.1. Introduction

Data from the summer Head Start program present a case of considerable practical importance. At stake is a question about a causal effect: whether Head Start participation improves test scores and subsequent school performance. Head Start was evaluated in 1969 by V. Cicirelli et al. [15], who argued that the summer Head Start program had no effect on achievement or ability test scores. The data considered included socio-economic status variables for Head Start participants and non-participants matched by race and neighborhood, and scores on two tests, the Illinois Test of Psycholinguistic Abilities and Metropolitan Readiness Test. Unfortunately, the tests were administered to participants only after participation in summer Head Start. J. Magidson [69], employing a causal model to make his case, took issue with a portion of the Cicirelli findings, and argued that summer Head Start participation did have a positive effect on test scores. Magidson's conclusion was criticized by Bentler and Woodward [4], principally on the grounds that the effect was not statistically significant. They also suggested, but did not analyze, a different causal structure for the data. Pursuing the suggestion, Magidson and Sorbom argued that a second model also led to the conclusion that the program had the desired effect, but again the positive effect was not statistically significant. The Magidson-Sorbom model has been described again by Sorbom [97], and yet again by Joreskog and Sorbom in the LISREL VI manual.

The variables in Magidson and Sorbom's [70] model for data from white 6 year old participants and non-participants in summer Head Start, with their abbreviations, are:

1. Mother's education = me
2. Father's education = fe
3. Father's occupation = fo
4. Family income = i
5. The child's score on the first of two tests given to Head Start participants at the end of the summer program. = t1
6. The child's score on the second test = t2

The sample correlations among these variables are:

Participants (N = 148)

me	1.00					
fe	.466	1.00				
fo	.253	.203	1.00			
in	.361	.182	.377	1.00		
t1	.275	.265	.208	.084	1.00	
t2	.256	.122	.251	.198	.664	1.00

Non-participants (N = 155)

me	1.00					
fe	.561	1.00				
fo	.224	.342	1.00			
i	.306	.215	.387	1.00		
t1	.239	.215	.196	.115	1.00	
t2	.281	.297	.234	.162	.635	1.00

Figure 8-43: Head Start Correlation Data

The strategy Magidson and Sorbom pursue is to treat the participants and non-participants as separate groups to which the same causal model applies, and to assume that all linear coefficients, save one associated with the effect of Head Start participation on cognitive ability, have the same values in the two groups. They assume that cognitive ability is a linear function of SES and of Head Start participation:

$$CA = a P + b SES + E$$

where P is a dummy variable equal to zero for non-participants and equal to 1 for participants. The object is to estimate the coefficient a. They use LISREL V, while our statistical analyses are done with EQS and LISREL IV.

Magidson and Sorbom's initial model for the SES indicators is pictured in Fig. 8-44.

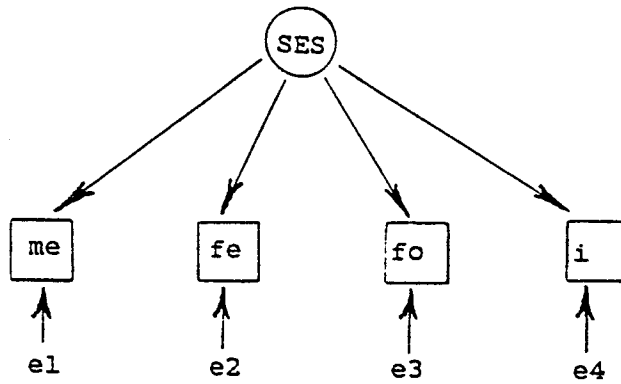


Figure 8-44: Magidson and Sorbom's Initial SES Model

The fit for this model is unacceptable (chi square with 10 d.o.f. is 43.6). The LISREL recommendation for revision of the model is to free a parameter for correlated error between *me* and *fe*, giving the model shown in Fig. 8-45.

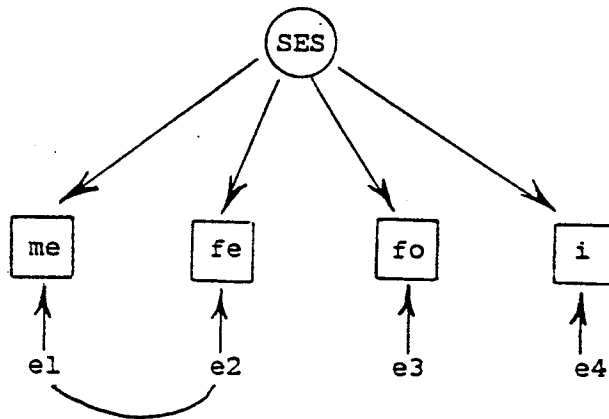


Figure 8-45: LISREL Suggestion

Here chi square is 31.04 with 22 d.o.f. ($p = .10$). On the pooled data, however this model gives a chi square with $p = 0$. This is not a happy result, since the fundamental assumption is that the submodel applies to the subgroups individually and jointly.

Magidson and Sorbom postulate that post-test scores depend on latent cognitive ability, which in turn depends on socio-economic status, and perhaps on Head Start participation. Their full model for each group is shown in Fig. 8-46.

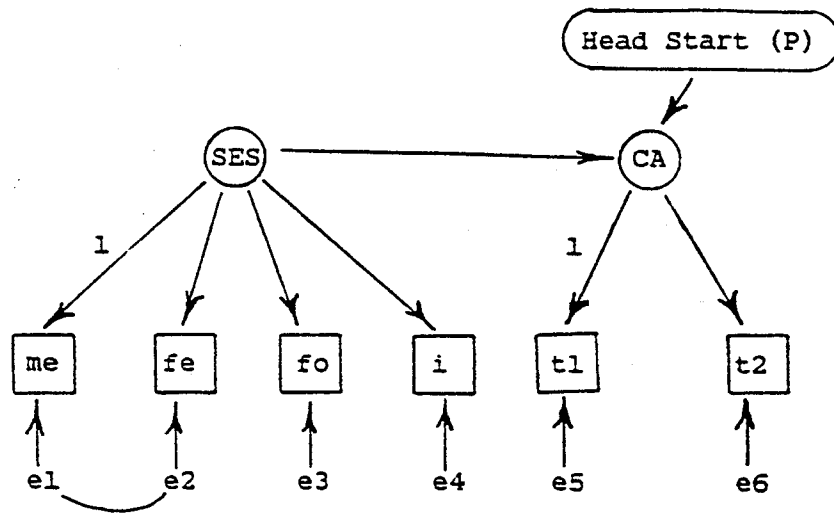


Figure 8-46:

Ignoring for the moment the Head Start dummy variable, and with all parameters except for the SES \rightarrow CA structural coefficient specified to be the same in the two groups, chi square is 31.08 with 23 degrees of freedom, a statistic whose probability is about .10. But if correct, Magidson and Sorbom's revised model should apply to the participant and non-participant groups separately. Using EQS we find for the participant group alone that the chi square value (15.38 with 7 degrees of freedom) has a probability of .03. In view of the small sample size, this is not reassuring.

8.7.2. Difficulties with the Magidson-Sorbom Model

There are several difficulties with the Magidson-Sorbom model. First, the fit is unacceptable, both for the socio-economic status model and for the model as a whole on the participant data. Poor fit is not a sufficient reason to reject a model if there are compelling reasons to think that the general assumptions are correct, and if no plausible alternatives provide better fit. But in this case we have no powerful reasons to support the model's general assumptions, the sample size is not large, and we have no reason to believe that alternative models might not give much better fit. A second difficulty is the assumption that the structural parameters are constant across the two groups. That hypothesis is not tested in Magidson and Sorbom's analysis, but it could be given a rough test by estimating the coefficients separately on the two groups, and comparing them. But

there is little point in carrying out such a comparison until one has a common model that fits each group separately.

There are two further substantive difficulties. Magidson and Sorbom's model claims that occupation has no direct effect on income, and that the correlation of the two is only obtained through a common socio-economic status variable. This hypothesis does not seem likely to be a reasonable approximation to the truth. Finally, the Magidson Sorbom model assumes that family socio-economic status causes mothers education, father's education, father's occupation and family income. It is not clear that this assumption makes sense. For reasons described in an earlier chapter, family socio-economic status ought, if anything, to be an effect of the these measured variables, not a common cause of them. The model assumes that parental education, occupation and income have no effect on children's cognitive ability; the common socio-economic status variable alone affects cognitive ability. This seems substantially less likely than that parental education, occupation and income, no doubt through their effects on unmeasured intervening variables, have an effect on children's cognitive ability.

8.7.3. The TETRAD Analysis

In investigating alternative models for this case, we will consider three different initial hypotheses. We will first consider the initial model proposed by Magidson and Sorbom, which does not contain the correlated errors for fe and me . Second, we will consider an initial model obtained by adding an $fo \rightarrow i$ connection to Magidson and Sorbom's model. Third, we will investigate the hypothesis that the measured socio-economic status variables affect some latent variable, which might be interpreted as cognitive ability, or as academic test taking ability, and that latent variable in turn affects test performances. We find a number of models that have conventionally acceptable fit for both participant and non-participant groups. Since we do not carry out a full search, there are quite possibly several other models that have equally good fit, explanatory power and simplicity. We use the participant data with TETRAD to search for good models, and test the models obtained both on the non-participant and on the participant data. Contrary to conventional wisdom that prohibits testing a model on the same data used to discover it, on the grounds that such tests are "tautological," the participant data sometimes provides a more severe test of the models than does the non-participant data, producing much lower p values.

Magidson and Sorbom's Initial Model

When we pass the initial model through TETRAD and ask for the suggested sets, we obtain the following:

Sets of suggested treks at significance level = 0.5156

{fo-t1 }

Sets of suggested treks at significance level = 0.6936

{me-fo fe-t1 fo-t1 }

Sets of suggested treks at significance level = 0.7720

{me-fe me-fo fe-t1 fo-t1 i-t1 }
 {me-fe me-fo me-t1 fe-t1 fo-t1 }
 {me-fe me-fo me-t1 fe-t1 i-t1 }
 {me-fe me-fo me-t1 fo-t1 i-t1 }

Figure 8-47:

We note that the program does not suggest the fo--i trek which we consider to be required on substantive grounds.

We further note that the second unit of suggested treks could be simply realized by a common cause of fo, t1 and fe, if a trek between fe and fo were permissible. Equally, a common cause of t1, me and fo could be implemented if a trek between me and t1 were permissible. The third unit of suggested sets does not contain a trek between fe and fo, but it does (in the second set) contain a trek between me and t1. Moreover, the second suggested set in the third unit contains the unique suggested set of the second unit. Both our heuristics and simplicity recommend implementing the second suggested set of trek additions in the third unit. We realize the recommended treks by correlating the errors of me and fe, and by introducing a new common cause of me, fo, and t1. The graph of the model is:

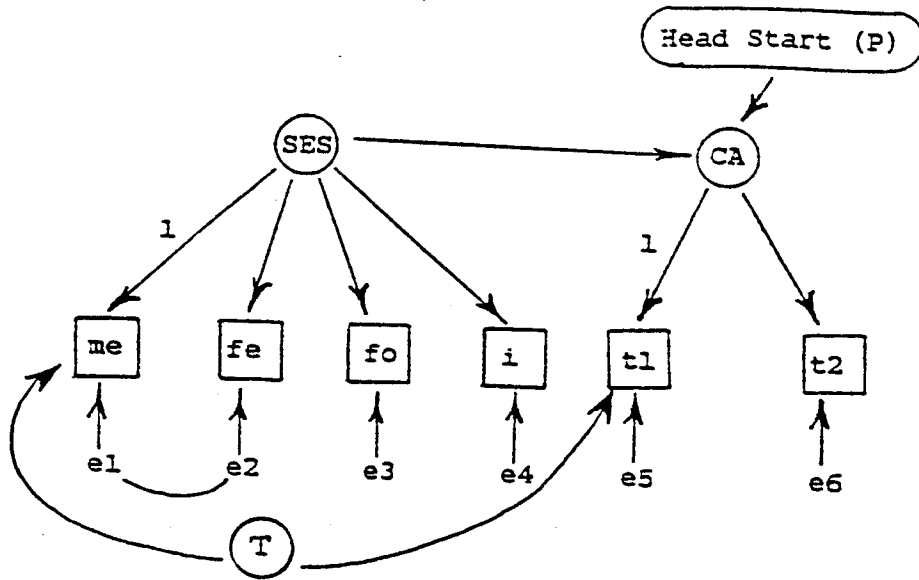


Figure 8-48: A Revision Suggested by TETRAD

This model has a low TTR value, .1415. The initial model has a TTR of 1.34.

Occupation and Income

We now assume Magidson and Sorbom's initial model, but with the addition of fo → i. If we give this model to TETRAD and ask for suggested sets, TETRAD ignores the fo → i connection (because the model is not skeletal), and we get the same suggestions as before. We consider instead the RTTR chart. We choose a significance level of .7, roughly the level of the second unit of suggested sets of trek additions. The RTTR chart is given below:

The graph analyzed in this example is:

fo→i SS→me SS→fe SS→fo SS→i SS→CA CA→t1 CA→t2

The significance level is: 0.7000

The sample size is: 148

Edge	Rttr	D(I-H)	I(H-I)
me→ fe	0.383	7	1
fe→ me	0.383	7	1
me C fe	0.383	7	1
me→ fo	0.423	10	1
fo→ me	0.423	10	1
me C fo	0.423	10	1
me→ i	0.275	6	1

i-> me	0.423	10	1
me C i	0.275	6	1
me-> t1	0.192	6	1
t1-> me	0.326	9	2
me C t1	0.192	6	1
me-> t2	0.217	6	1
t2-> me	0.326	9	2
me C t2	0.217	6	1
fe-> fo	0.297	8	3
fo-> fe	0.297	8	3
fe C fo	0.297	8	3
fe-> i	0.158	4	3
i-> fe	0.297	8	3
fe C i	0.158	4	3
fe-> t1	0.261	7	0
t1-> fe	0.419	10	1
fe C t1	0.261	7	0
fe-> t2	0.276	6	1
t2-> fe	0.419	10	1
fe C t2	0.276	6	1
fo-> t1	0.365	10	1
t1-> fo	0.656	15	2
fo C t1	0.365	10	1
fo-> t2	0.443	10	1
t2-> fo	0.656	15	2
fo C t2	0.443	10	1
i-> t1	0.365	10	1
t1-> i	0.377	9	2
i C t1	0.237	6	1
i-> t2	0.443	10	1
t2-> i	0.377	9	2
i C t2	0.235	6	1

The Total Tetrad Residual, or TTR, is: 0.79826

We eliminate all directed edges from test score variables to indicators of SES because they make no sense. The Rtrr chart suggests several possible additions:

i→ t2	0.443	10	1
fo→ t2	0.443	10	1
fo C t2	0.443	10	1
fe C t1	0.261	7	0
fe→ t1	0.261	7	0
me→ fo	0.423	10	1
fo→ me	0.423	10	1
me C fo	0.423	10	1

Figure 8-49: The Rtrr Chart's Best Candidates

A proper search would now treat each of these edges as second level nodes in a search tree, and investigate third level nodes.²⁹ We will not pursue the search, however, but we will subsequently give chi square results for these models.

An Alternative Hypothesis

The Magidson-Sorbom model assumes that parental education, occupation and income are *not* causes of whatever abilities account for performance on the two tests. It seems to us more plausible that these measured variables do cause the capacities exhibited on the tests, and we will use TETRAD to explore that hypothesis. Since parental education, occupation and income are highly correlated we assume that they have a common unmeasured source of covariance, but we give no interpretation to this latent variable.

The initial model we assume looks as follows (neglecting error terms):

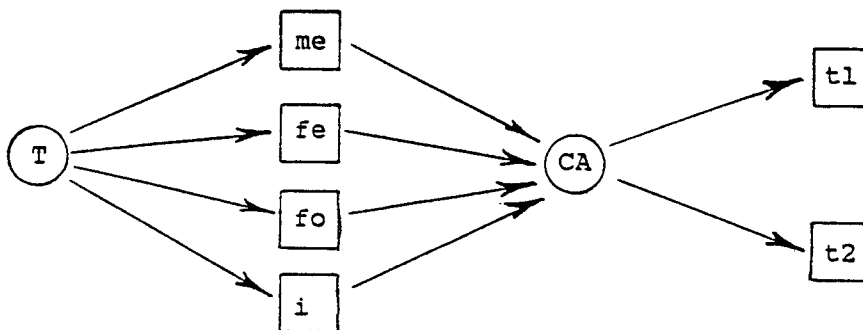


Figure 8-50: An Alternative Causal Skeleton

²⁹We note in passing that the correlated error between me and fe, suggested by LISREL for addition to the Magidson and Sorbom initial model also has the highest Rtrr value for that model. But when the causal connection between occupation and income is imposed, the Rtrr chart shows that an me fe connection is no longer distinguished either by Rtrr value or by I(H-I).

This model is not a conventional multiple indicator model, because it has directed edges from measured variables to latent variables. TETRAD will nonetheless suggest trek additions for it if we request them. The suggested treks are supposed to be those that will not prevent the implication of any tetrad equations that hold at the appropriate significance level. A rough but quick check of whether the TETRAD heuristic is working as it should can be obtained from the Rtrr chart at the same significance level: every trek in any suggested set should have a $I(H-I)$ value of 0. When we ask TETRAD to suggest trek additions to the initial model we obtain the following:

Sets of suggested treks at significance level = 0.0367

```
{me-fe }
{fo-i }
```

Sets of suggested treks at significance level = 0.0927

```
{me-fe fe-t1 }
{me-fe fe-t2 }
{fe-t1 fo-i }
{fe-t2 fo-i }
```

Sets of suggested treks at significance level = 0.2655

```
{me-fe fe-t1 i-t1 }
{me-fe fe-t1 i-t2 }
{me-fe fe-t2 i-t1 }
{me-fe fe-t2 i-t2 }
{fe-t1 fo-i i-t1 }
{fe-t1 fo-i i-t2 }
{fe-t2 fo-i i-t1 }
{fe-t2 fo-i i-t2 }
```

Figure 8-51:

The Rtrr chart with the significance level set at .1 shows that the first unit of suggested sets does have $I(H-I)$ equal to zero, but that the suggested trek additions in higher units are not reliable.

Of the suggested trek additions in the first unit, we prefer $fo \rightarrow i$, for reasons already explained. If we add $fo \rightarrow i$ to the initial model and again ask for the Rtrr chart at the same significance level we obtain:

The graph analyzed in this example is:

me->CA fe->CA fo->in fo->CA in->CA T ->me T ->fe T ->fo T ->in CA->t1 CA->t2

The significance level is: 0.1000

The sample size is: 148

Edge	Rttr	D(I-H)	D(H-I)
me-> fo	0.027	0	1
fo-> me	0.027	0	1
fo C me	0.027	0	1
me-> in	0.027	0	1
in-> me	0.027	0	1
in C me	0.027	0	1
me-> t1	0.186	3	3
t1-> me	0.110	1	3
t1 C me	0.083	1	2
me-> t2	0.186	3	3
t2-> me	0.110	1	3
t2 C me	0.083	1	2
CA-> me	0.027	0	1
fe-> fo	0.027	0	1
fo-> fe	0.027	0	1
fo C fe	0.027	0	1
fe-> in	0.027	0	1
in-> fe	0.027	0	1
in C fe	0.027	0	1
fe-> t1	0.186	3	3
t1-> fe	0.145	3	1
t1 C fe	0.118	3	0
fe-> t2	0.186	3	3
t2-> fe	0.145	3	1
t2 C fe	0.118	3	0
CA-> fe	0.027	0	1
fo-> t1	0.186	3	3
t1-> fo	0.179	2	4
t1 C fo	0.152	2	3
fo-> t2	0.186	3	3
t2-> fo	0.179	2	4
t2 C fo	0.152	2	3
CA-> fo	0.027	0	1
in-> t1	0.186	3	3
t1-> in	0.123	1	3
t1 C in	0.095	1	2

in-> t2	0.186	3	3
t2-> in	0.123	1	3
t2 C in	0.095	1	2
CA-> in	0.027	0	1
T -> t1	0.186	3	3
T -> t2	0.186	3	3
CA-> T	0.027	0	1

The Rtrr chart shows that there are no outstanding elaborations at this stage in the search. A number of edges will reduce the TTR by .186 but cost three equations in explanatory power, and a number will reduce the Rtrr by .145 and only cost 1 equation in explanatory power.

A thorough search would therefore generate a great many nodes at this level of the search tree. We will not conduct a thorough search, but we will consider the statistical properties of the model that adds the further edge fe->t1. For symmetry, we also consider the effects of adding fe->t1 and fe->t2 to our skeleton and fo -> i.

8.7.4. Goodness of Fit Results

The following table summarizes the chi square results for the models we have considered.

Model	Data Set	DF	X ²	p(X ²)
Magidson & Sorbom's Skeleton				
1p) me C fe	Part.	7	15.38	.031
1np) me C fe	Non-part.	7	11.79	.107
2p) me c fe, T->me,fo,t1	Part.	4	11.05	.026
2np) "	Non-part.	4	4.35	.360
3p) fo->i, i->t2	Part.	6	10.09	.121
3np) fo->i, i->t2	Non-part.	6	8.39	.211
4p) fo->i, fo->t2	Part.	6	11.66	.070
4np) fo->i, fo->t2	Non-part.	6	8.12	.230
5p) fo->i, fo C t2	Part.	6	11.66	.070
5np) fo->i, fo C t2	Non-part.	6	8.12	.230
6p) fo->i, fe C t1	Part.	6	9.12	.167
6np) fo->i, fe C t1	Non-part.	6	8.12	.236
7p) fo->i, fe->t1	Part.	6	9.12	.167
7np) fo->i, fe->t1	Non-part.	6	8.12	.236
8p) fo->i, me->fo	Part.	6	57.82	<.001
8np) fo->i, me->fo	Non-part.	6	70.85	<.001
9p) fo->i, fo->me	Part.	6	11.93	.064
9np) fo->i, fo->me	Non-part.	6	73.35	<.001
10p) me C fo, fo->i	Part.	6	11.93	.064
10np) "	Non-part.	6	3.11	.795
Our Skeleton				
11p) fo -> i	Part.	4	10.22	.037
11np) "	Non-part.	4	6.83	.145
12p) fo->i, fe-t1	Part.	3	4.02	.260
12np) "	Non-part.	3	6.44	.092
13p) fo->i, fe->t1, fe->t2	Part.	2	4.02	.134
13np) "	Non-part.	2	6.44	.040

Figure 8-52: Goodness of Fit Results

These results were all obtained using the EQS program. The coefficient estimates are reasonably stable between groups for significant models, such as models 4 and 12.

8.7.5. Conclusion

Although we have not carried out a thorough search, the case does illustrate how TETRAD can be used in combination with substantive hypotheses about the domain. With the program models are found that fit the data for both participant and non-participants in summer Head Start.

8.8. ACHIEVEMENT, ABILITY, AND APPROVAL IN SCHOOL AGE CHILDREN

8.8.1. Introduction

This section is concerned with two points. First, every case so far examined in this book can be processed using the version of TETRAD on the floppy disk, which is limited to 9 variables. On the personal computer, the program can, however, process models with as many as 23 variables. We do not know the size limits for models processed on the Microvax version of the program, but they are considerably larger. This section illustrates the use of the program on a series of models each with 19 variables. All of the work was accomplished on a IBM compatible machine with 256K RAM and without an 8087 coprocessor. Processing models of this size can take as long as several hours, depending on the information requested. It is most conveniently done in batch mode, or while the user has something else to do that does not require the computer. The work described in this section took about an hour of user time and about a half a day of computer time.

Second, we have emphasized throughout the importance of searching for alternative models. But for the most part the cases we have considered have shown how TETRAD can be used to search for elaborations of an initial model, not how the program can be used to search for alternative initial models. The case we consider here is chosen to illustrate *one* way in which TETRAD can be used to search for alternative skeletal, or nearly skeletal, models. The models we found were not subjected to a chi square test, but we are confident that all of them would yield p values that are essentially zero. The search we describe is partial, but a more thorough search could be executed using the same techniques. We stop with a skeletal model that might be given to TETRAD for further elaboration. We do not investigate these elaborations here, but consider only the problem of locating plausible initial models.

8.8.2. The Maruyama and McGarvey Study

The data we consider are contained in a study by Maruyama and McGarvey, [72], of the causes of scholastic achievement. The data consist of correlations³⁰ of a number of measures of ability, socio-economic status, peer popularity, evaluations of parents and teachers, and academic achievement. The measures, with our abbreviations, are:

³⁰The data are in fact partial correlations, with sex and grade level controlled for.

1. Duncan Socio-Economic Index = se
2. Education of head of household = ed
3. Number of rooms per person in house = rm
4. Scores on standardized verbal tests = va
5. Verbal grades = vg
6. Peabody Picture Vocabulary Test = pv
7. Raven's Progressive Matrices = pm
8. Playground popularity = pp
9. Seating popularity = sp
10. Schoolwork popularity = wp
11. Father's evaluation = fe
12. Mother's evaluation = me
13. Teacher's evaluation = te

Maruyama and McGarvey assume these variables form five clusters, with the members of each cluster sharing a common latent variable. The latent variables are named Socio-economic status (SS), Academic ability (AB), Achievement (AC), Acceptance by peers (AP), and Acceptance by significant adults (AS).

The model Maruyama and McGarvey consider is given by the following graph, in which we have omitted error terms.

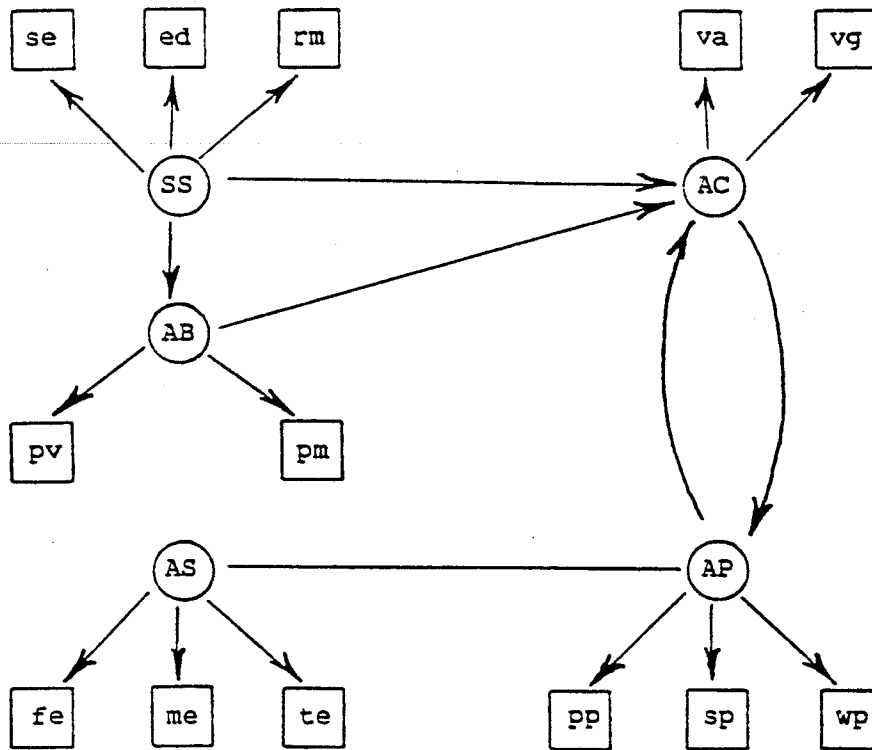


Figure 8-53: Maruyama and McGarvey Model

Maruyama and McGarvey are concerned with whether academic achievement causes "acceptance by significant others," or whether instead acceptance causes achievement, or whether each has an effect on the other, or whether, finally, any covariation between acceptance and achievement is due to common causes. The model depicted above is proposed "to test the competing alternatives." In effect, Maruyama and McGarvey simply use LISREL to estimate the standardized coefficients associated with $AC \rightarrow AP$ and $AP \rightarrow AC$. They find that the first is .270 and the second is .232, and while they perform no formal test of significance of the difference, it seems likely that the difference is not significant. The overall model yields a chi square statistic with a p value of zero.

Their study is offered to illustrate a strategy for using LISREL to decide causal questions. There are any number of reasons for rejecting the procedure, and for denying that the particular study is of any weight. The procedure does not really test the alternative causal relations between acceptance or approval and academic achievement. Maruyama and McGarvey note that there are a number of other alternative plausible connections among the latent variables, and that their model excludes the very real possibility that acceptance and achievement have common causes. Their strategy is in effect to pick a particular causal model from among many alternative possibilities, and to have edges in the model that

represent particular causal hypotheses of interest, and use estimates of these coefficients to decide among the hypotheses. Without persuasive justification for the model assumed, this seems a risky strategy. The particular case is still more dubious. Maruyama and McGarvey obtain an insignificant difference from a model that fails a chi square test on a sample of middling size.³¹

Before we begin our search, we note that TETRAD makes it easy to apply another strategy. Citing Bentler and Woodward, Maruyama and McGarvey suggest that one might compare the two models obtained from theirs by deleting, respectively, the AP→AC edge and the AC→AP edge. An approximate comparison of these alternatives can be obtained in one step with TETRAD by omitting *both* edges from the model, and asking for the Rtrr chart (for latent-latent connections only) of the submodel that results. The output is obtained in about ten minutes.

Edge	Rtrr	D(I-H)	I(H-I)
AC→AP	16.295	247	179
AP→AC	13.536	203	133

AC→AP does substantially more to reduce the TTR value than does AP→AC, and we would therefore expect the deletion of AC→AP from the Maruyama and McGarvey model to give poorer fit than the deletion of the AP→AC edge.

8.8.3. A TETRAD Search

Maruyama and McGarvey's study presents a case in which there are thirteen variables clustered into five groups. There are an enormous number of ways of specifying causal connections among the five latent variables (in fact, 4^{10} different ways if no further variables are introduced), and we have few strong theoretical guides to exclude possibilities. We can be confident that peer and adult approval do not cause socio-economic status, and neither does achievement. We can be less confident that achievement does not cause ability, and we have some evidence (see Brophy and Good, [12]) that socio-economic status has an effect on adult approval, but we cannot be confident of much else *a priori*. When we have limited knowledge of this kind, the best we can do is to search for

³¹Equally dubious, the data are partialled on a dichotomous variable (sex) and then the partial correlations are assumed to be generated by normally distributed random variables.

alternatives that give a reasonably good fit, explain patterns in the data, and are not excluded by what we think we know.

The actual problem is rather worse, since alternative clusterings are surely possible. We will, however, assume for purposes of illustration that the measured variables are clustered in the way Maruyama and McGarvey propose, and consider how we might search through the possible causal connections among the latent variables.

Our procedure is to begin with a model in which there are *no* causal connections among the latent variables. Because there are no treks connecting variable pairs in most foursomes of variables, the disconnected cluster model will imply a great many tetrad equations and have an enormous TTR value. We give this model to TETRAD and ask for the Rtrr chart for edges connecting latent variables. We add the connection that has the highest Rtrr value, and run the modified model through TETRAD again. We continue in this way until there are edges from measured to measured variables, or from latent to measured variables, that have Rtrr values greater than any addition connecting only latent variables. If at any stage we find that several alternative additions produce nearly the same Rtrr value, the search branches, and every alternative is considered. Since the primary problem is to find latent variable models that dramatically reduce the TTR of the disconnected clusters, considerations about the explanation of tetrad equations plays a secondary role.

This search procedure is not very thorough, and it will undoubtedly overlook many excellent models. It is undoubtedly a far better search, however, than people are able to do without the aid of the program. We will not even carry out the full search procedure, but will rather follow only one line, noting where the search might reasonably branch.

We note at the outset that the Maruyama and McGarvey model has a TTR value of 12.9. Their model also implies some vanishing partial correlations, since it implies that the indicators of adult approval are uncorrelated with the indications of socio-economic status and ability. At a significance level of .4, the model explains 52 vanishing partial correlations that are not rejected by the significance test, and implies 38 vanishing partial correlations that do not hold. Another 165 vanishing partial correlations hold at that significance level, but are not implied by the model. The tetrad residuals are in this case more powerful indicators of fit. There are 1031 tetrad equations that hold at the .4 significance level, and the model implies 240 that do not hold and 499 that do hold. We will consider only the TTR values in our search.

We begin our search with the model in which the five clusters are disconnected, that is:

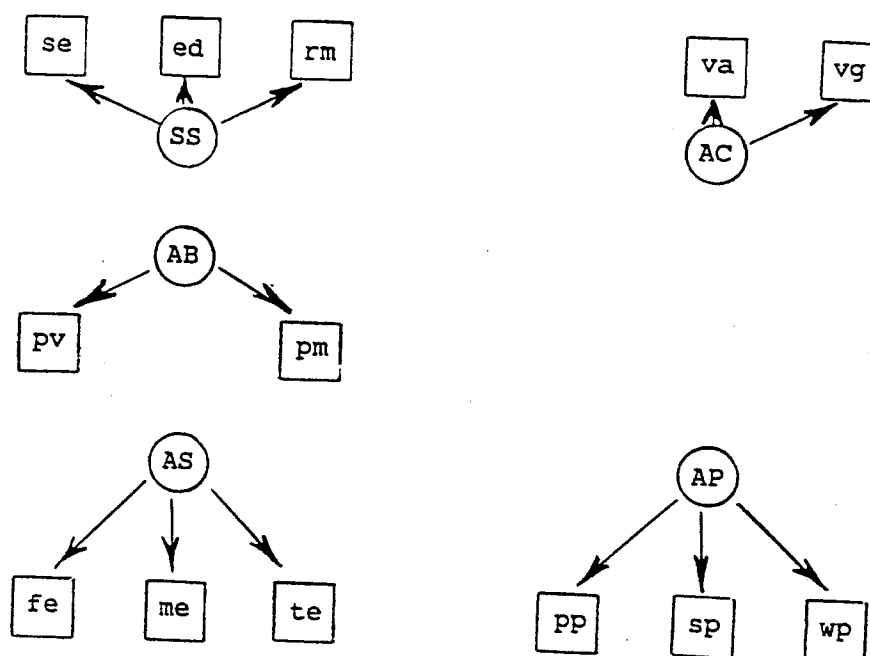


Figure 8-54: Disconnected Model

We give this model to TETRAD and ask for the Rtrr chart. The connections among latent variables have the largest Rtrr values, and of these, the connections between SS and AB most reduce the TTR value. In this case it makes no difference for any connection which way the arrows go, or whether there is a common cause. The relevant output is given below:

The graph analyzed in this example is:

AC->va AC->vg SS->rm SS->se SS->ed AB->pm AB->pv AP->sp AP->pp
AP->wp AS->fe AS->me AS->te

The significance level is: 0.4000

The sample size is: 249

Edge	Rttr	D(I-H)	I(H-I)
SS->AC	5.360	68	16
SS->AB	4.711	55	29
SS->AP	2.070	35	55
SS->AS	2.565	37	53
AC->AB	6.331	67	5
AB->AC	6.331	67	5
AC->AP	5.338	76	8
AP->AC	5.338	76	8
AC->AS	4.438	54	30
AS->AC	4.438	54	30
AB->AP	2.708	54	30
AP->AB	2.708	54	30
AB->AS	3.232	57	27
AS->AB	3.232	57	27
AP->AS	3.244	56	34
AS->AP	3.244	56	34

The Total Tetrad Residual, or TTR, is: 64.05317

I-H, the total number of equations implied by the model that do not hold in the data, is: 1024

H-I, the total number of equations that hold in the data but are not implied by the model, is: 2

The number of equations that hold in the data is 1031

The number of equations explained by the model is 1029

The Total Partial Residual, or TPR, is: 75.67931

PI-PH, the total number of equations implied by the model that do not hold in the data, is: 523

PH-PI, the total number of equations that hold in the data but are not implied by the model, is: 3

The number of equations that hold in the data is 217

The number of equations explained by the model is 214

Evidently a connection between ability and achievement is most important.

We add $AB \rightarrow AC$ to the disconnected model, since it seems the most plausible connection between the two variables.

When we give the revised model to TETRAD and ask for the Rtrr chart we find that connections between AB and SS have the highest Rtrr value, 9.7. We can introduce a connection between AB and SS by means of a new latent variable, T, which serves as a common cause. One could as well introduce a direct effect of SS on AB. We obtain two alternative revised models, which look like this:

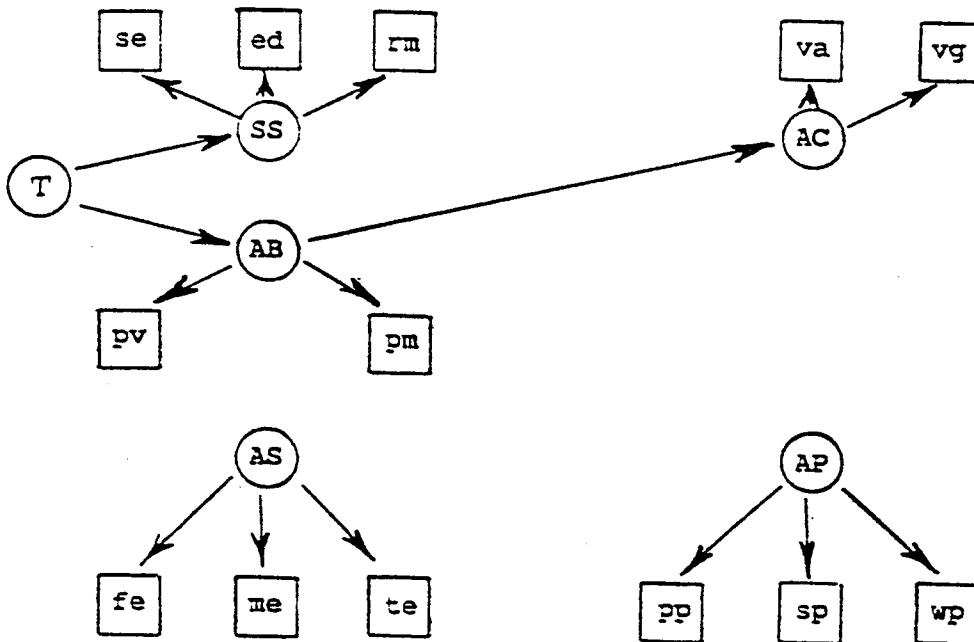


Figure 8-55: Disconnected Model plus $AB \rightarrow AC$, $T \rightarrow AB$, $T \rightarrow SS$

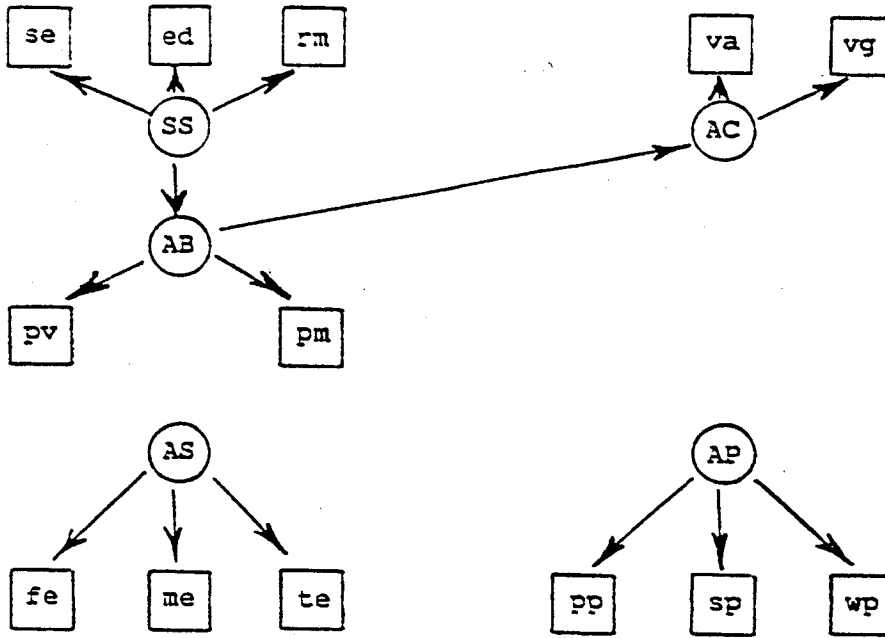


Figure 8-56: Disconnected Model plus AB→AC, SS→AB

We will pursue only the first branch of the search. When we again give the model of Figure 8-55 to the program the following lines of the Rtrr chart give the best results.

The graph analyzed in this example is:

AC→va AC→vg SS→rm SS→se SS→ed AB→pm AB→pv AB→AC AP→sp
 AP→pp AP→wp AS→fe AS→me AS→te T→SS T→AB

The significance level is: 0.4000

The sample size is: 249

Edge	Rtrr	D(I-H)	I(H-I)
AC→AP	10.771	171	111
AC→AS	10.610	163	119
T→AP	10.895	170	148
T→AS	11.533	176	142

T→AS has the highest Rtrr value, 11.533. This is a plausible connection, supported by research cited by Brophy and Good [12]. The other edges shown are close, so again the search might branch, but we will follow only the line indicated by this alternative. We add T→AS to the model, obtaining:

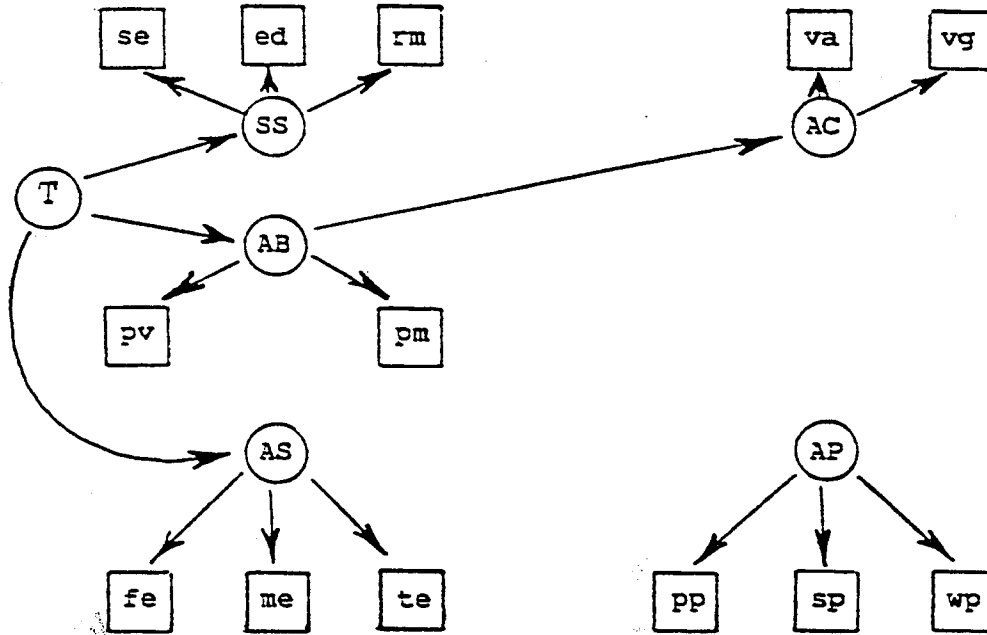


Figure 8-57: Disconnected Model plus AB→AC, T→AB, T→SS, T→AS

Continuing with the search, we give this model to TETRAD and request another Rtrr chart. We show the relevant parts of it below.

The graph analyzed in this example is:

AC->va AC->vg SS->rm SS->se SS->ed AB->pm AB->pv AB->AC AP->sp AP->pp
 AP->wp AS->fe AS->me AS->te T->SS T->AB T->AS

The significance level is: 0.4000

The sample size is: 249

Edge	Rttr	D(I-H)	I(H-I)
AC->AP	19.257	344	280
AP->AC	13.064	212	172
AB->AP	16.256	296	220
AP->AB	17.066	303	285
AP->AS	14.902	280	224
AS->AP	16.019	288	282
T->AP	16.072	260	220

Here there is a unique best edge, AC->AP. So far we have the following model.

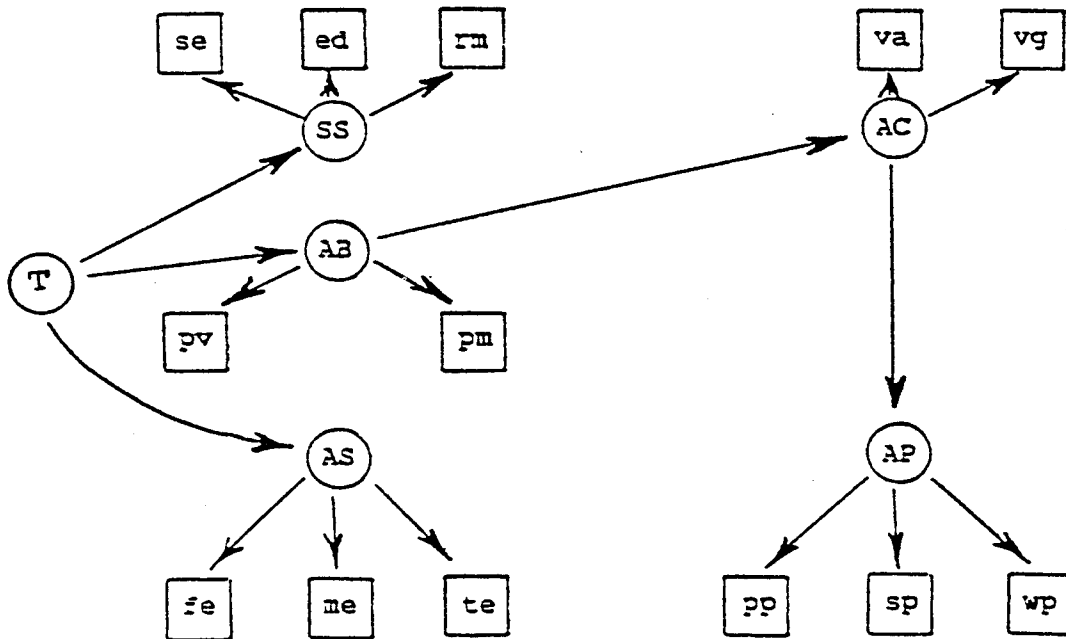


Figure 8-58: Disconnected Model plus AB->AC, T->AB, T->SS, T->AS, AC->AP

Giving this model to TETRAD, we request an Rtrr chart and standard results. The relevant output is shown below.

The graph analyzed in this example is:

AC->va AC->vg AC->AP SS->rm SS->se SS->ed AB->pm AB->pv AB->AC
 AP->sp AP->pp AP->wp AS->fe AS->me AS->te T->SS T->AB T->AS

The significance level is: 0.4000

The sample size is: 249

Edge	Rttr	D(I-H)	I(H-I)
AC->SS	2.628	54	66
SS->AC	3.817	75	75
AC->AB	1.863	33	27
AP->AC	1.306	19	29
AC->AS	3.143	63	57
AS->AC	3.817	75	75
AC->T	1.863	33	27
T->AC	1.863	33	27
SS->AP	5.483	102	162
AP->SS	4.438	87	141
AB->AP	1.306	19	29
AP->AB	4.474	74	106
AP->AS	5.462	110	118
AS->AP	5.483	102	162
AP->T	4.474	74	106
T->AP	3.537	58	98

The Total Tetrad Residual, or TTR, is: 17.24747

I-H, the total number of equations implied by the model that do not hold in the data, is: 321

H-I, the total number of equations that hold in the data but are not implied by the model, is: 469

The number of equations that hold in the data is 1031

The number of equations explained by the model is 562

The Total Partial Residual, or TPR, is: 0.00000

PI-PH, the total number of equations implied by the model that do not hold in the data, is: 0

PH-PI, the total number of equations that hold in the data but are not implied by the model, is: 0

The number of equations that hold in the data is 0

The number of equations explained by the model is 0

Notice that now the model implies no vanishing partial constraints, because it is now trek connected. The model we have now reached reduces the TTR to around 17.2, and it still explains 562 equations that hold at a significance level of .4. The Rtrr chart contains three edges that look most promising, SS→AP, AP→AS, and AS→AP. That SS causes peer acceptance is certainly plausible, so we elaborate our model accordingly.

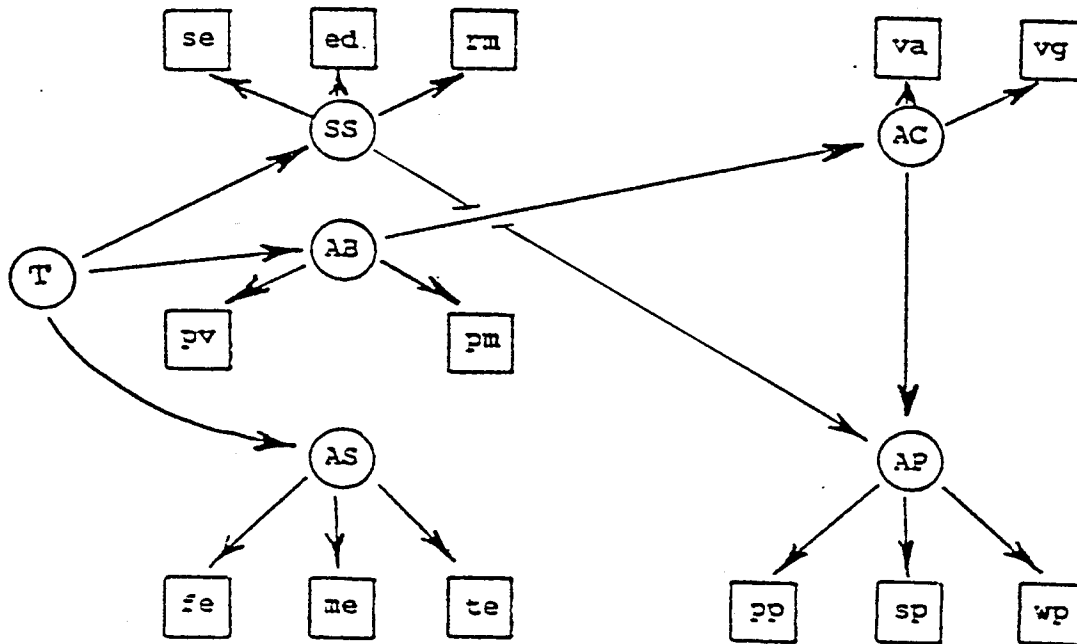


Figure 8-59: Disconnected Model plus AB→AC, T→AB, T→SS, T→AS, AC→AP, SS → AP

Running this model through TETRAD shows that the Rtrr chart now contains other edges that do more to reduce the TTR than any edge from a latent to another latent variable. Its performance on TETRAD's standard measures is shown below.

The graph analyzed in this example is:

AC->va AC->vg AC->AP SS->rm SS->se SS->ed SS->AP AB->pm AB->pv
 AB->AC AP->sp AP->pp AP->wp AS->fe AS->me AS->te T->SS T->AB T->AS

The Total Tetrad-Residual, or TTR, is: 11.76464

I-H, the total number of equations implied by the model that do not hold in the data, is: 219

H-I, the total number of equations that hold in the data but are not implied by the model, is: 631

The number of equations that hold in the data is 1031

The number of equations explained by the model is 400

Compare this model with Maruyama and McGarvey's. The new model is as plausible, and it has a TTR value of 11.7 compared to the Maruyama and McGarvey model TTR of 12.9. The new model implies no vanishing partial correlations that fail a significance test at a significance value of .4, whereas Maruyama and McGarvey's model implies 38 incorrect vanishing partial correlations. We therefore expect that the new model will provide a better fit, although we would still expect it to fail a chi square test at the .05 significance level.

We would not, on the basis of this data and this cursory search put any weight on the substantive claims our model makes. We remind the reader that any number of equally good models might be found by making different choices at stages in the search.

8.9. THE STABILITY OF ALIENATION

8.9.1. Introduction

A study by Wheaton et al., [107], on the stability of alienation has become a standard example in manuals for computer programs that perform statistical analyses of structural equation models. Joreskog and Sorbom [50] discuss the example in the LISREL manuals, Bentler discusses the example in the EQS manual, and we introduce it here to illustrate the use of the TETRAD program. Covariance data for this case are on the floppy disk. The file name is "Alien.dat"

The alienation study shows the use of both structural criteria and plausibility criteria in respecifying an initial causal model. The interpretation of variables in the initial model rules out a host of possible elaborations, and therefore reduces the search space significantly.

The initial causal model is pictured in Figure 8-59. The probability of its chi square statistic (with 6 degrees of freedom) is less than .001.

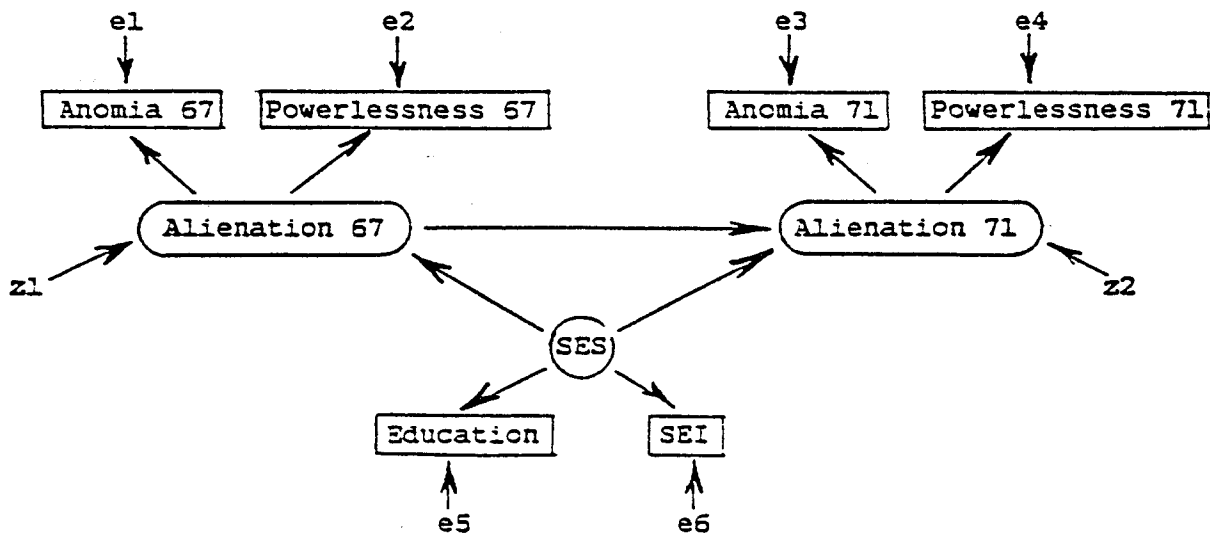


Figure 8-12: Alienation: Original Model

Alienation is a latent construct measured by anomia and powerlessness at two different times. SES is socio-economic status, also latent, measured by Duncan's socio-economic index and an index of educational achievement. Using the LISREL search procedure, the

initial model is amended by freeing two parameters that correspond to correlations for error terms for the same indicator measured at different times. The revised model, shown in Figure 8-61, has 4 degrees of freedom and $p = .335$, $n = 932$. It is proposed by Joreskog and Sorbom, and by Bentler.

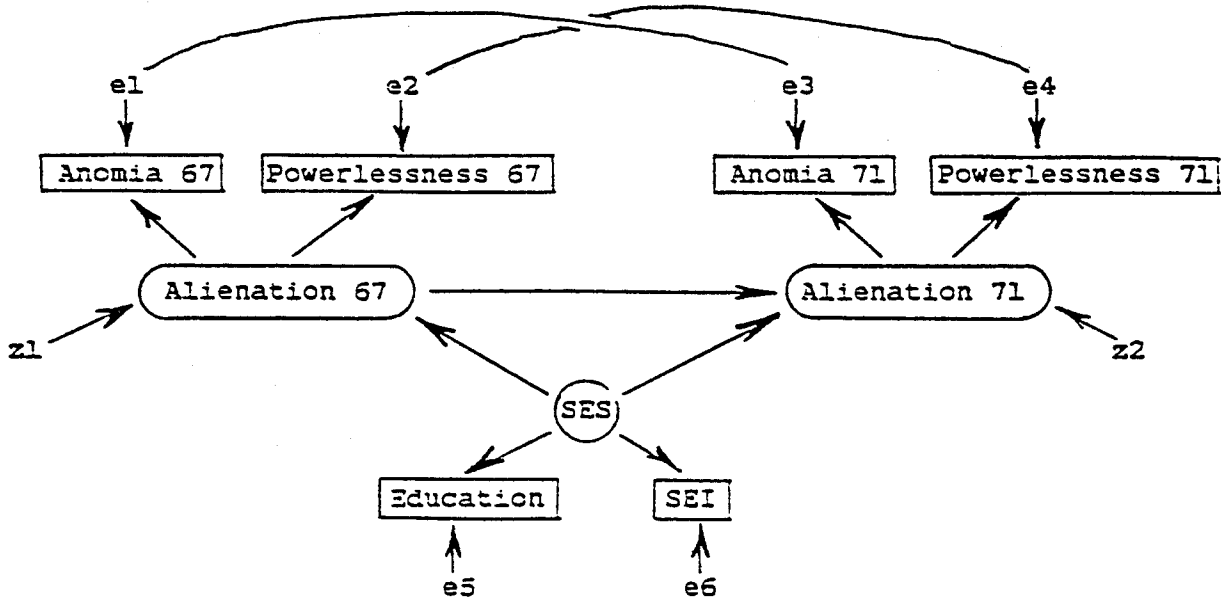


Figure 8-61: Alienation: Amended Model

It is certainly plausible that indicators measured at different times have correlated errors, and the probability value of the chi square statistic is high. But a researcher who settled for this account of the data would have missed many possibilities. We can find alternative models with the TETRAD program, and then use LISREL or EQS to compare the values of their chi square statistics.

8.9.2. The TETRAD Analysis

We use the abbreviation scheme given in Figure 8-62, and we give TETRAD the initial model pictured in Figure 8-63.

Alienation 67 = A6
 Alienation 71 = A7
 SES = SE
 Anomia 67 = a6
 Anomia 71 = a7
 Powerlessness 67 = p6
 Powerlessness 71 = p7
 SEI = se
 Education = ed

Figure 8-62: Abbreviation Scheme

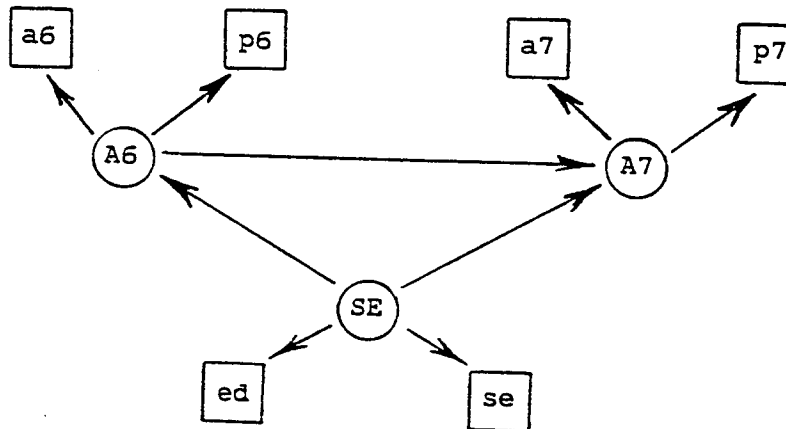


Figure 8-63: Alienation: Skeleton

We will conduct two searches for modifications of this initial model. One of them, using the Suggested Sets of trek additions, is easy and almost instantaneous, and leads to an extremely good model. The other search, using the Rtrr chart, is much more tedious, but it recovers the same model found with the Suggested Sets and another that is nearly as good.

Using the Suggested Sets

Asking for statistical information, we get the output we show below. Asking for three units of suggested sets, we get the results in Figure 8-64.

The graph analyzed in this example is:

A6->a6 A6->p6 A6->A7 A7->a7 A7->p7 SE->ed SE->se SE->A6 SE->A7

The significance level is: 0.0500

The sample size is: 932

Tetrad Equation	Residual	Impld.	Held	P(diff.)
a6 p6, a7 p7 = a6 a7, p6 p7	0.1510			0.0000
a6 p6, p7 a7 = a6 p7, p6 a7	0.2354			0.0000
a6 a7, p7 p6 = a6 p7, a7 p6	0.0844	y		0.0000
a6 p6, a7 ed = a6 a7, p6 ed	0.0014		y	0.9436
a6 p6, ed a7 = a6 ed, p6 a7	0.0618			0.0000
a6 a7, ed p6 = a6 ed, a7 p6	0.0604	y		0.0000
a6 p6, a7 se = a6 a7, p6 se	0.0287		y	0.1546
a6 p6, se a7 = a6 se, p6 a7	0.0501		y	0.0050
a6 a7, se p6 = a6 se, a7 p6	0.0214	y	y	0.1307
a6 p6, p7 ed = a6 p7, p6 ed	0.0638			0.0000
a6 p6, ed p7 = a6 ed, p6 p7	0.0570			0.0038
a6 p7, ed p6 = a6 ed, p7 p6	0.0068	y	y	0.6085
a6 p6, p7 se = a6 p7, p6 se	0.0572			0.0024
a6 p6, se p7 = a6 se, p6 p7	0.0289		y	0.1562
a6 p7, se p6 = a6 se, p7 p6	0.0283	y		0.0354
a6 p6, ed se = a6 ed, p6 se	0.2520			0.0000
a6 p6, se ed = a6 se, p6 ed	0.2334			0.0000
a6 ed, se p6 = a6 se, ed p6	0.0185	y		0.0447
a6 a7, p7 ed = a6 p7, a7 ed	0.0532	y		0.0000
a6 a7, ed p7 = a6 ed, a7 p7	0.0340		y	0.1044
a6 p7, ed a7 = a6 ed, p7 a7	0.0872			0.0000
a6 a7, p7 se = a6 p7, a7 se	0.0294	y		0.0320
a6 a7, se p7 = a6 se, a7 p7	0.0441			0.0362
a6 p7, se a7 = a6 se, p7 a7	0.0735			0.0000
a6 a7, ed se = a6 ed, a7 se	0.1981			0.0000
a6 a7, se ed = a6 se, a7 ed	0.1974			0.0000
a6 ed, se a7 = a6 se, ed a7	0.0007	y	y	0.9399
a6 p7, ed se = a6 ed, p7 se	0.1368			0.0000
a6 p7, se ed = a6 se, p7 ed	0.1266			0.0000
a6 ed, se p7 = a6 se, ed p7	0.0102	y	y	0.3616
p6 a7, p7 ed = p6 p7, a7 ed	0.0081	y	y	0.5296
p6 a7, ed p7 = p6 ed, a7 p7	0.1008			0.0000
p6 p7, ed a7 = p6 ed, p7 a7	0.0927			0.0000
p6 a7, p7 se = p6 p7, a7 se	0.0189	y	y	0.1534
p6 a7, se p7 = p6 se, a7 p7	0.0627			0.0000
p6 p7, se a7 = p6 se, p7 a7	0.0438			0.0304
p6 a7, ed se = p6 ed, a7 se	0.1351			0.0000
p6 a7, se ed = p6 se, a7 ed	0.1523			0.0000
p6 ed, se a7 = p6 se, ed a7	0.0172	y	y	0.1300
p6 p7, ed se = p6 ed, p7 se	0.1660			0.0000
p6 p7, se ed = p6 se, p7 ed	0.1735			0.0000
p6 ed, se p7 = p6 se, ed p7	0.0075	y	y	0.4903

a7 p7, ed se = a7 ed, p7 se	0.2638			0.0000
a7 p7, se ed = a7 se, p7 ed	0.2547			0.0000
a7 ed, se p7 = a7 se, ed p7	0.0091	y	y	0.2819

Sets of suggested treks at significance level = 0.1307

{a6-a7 }

Sets of suggested treks at significance level = 0.4903

{a6-a7 p6-se p7-se }

Sets of suggested treks at significance level = 0.5296

{p6-a7 p6-se p7-ed p7-se }
 {p6-a7 p6-se A6->a7 p7-ed }
 {p6-a7 p6-se A6->a7 p7-se }
 {p6-a7 p6-se p7-ed A6->p7 }
 {p6-a7 p6-se p7-se A6->p7 }
 {p6-a7 p6-se SE->a7 p7-ed }
 {p6-a7 p6-se SE->a7 p7-se }
 {p6-a7 p6-se p7-ed SE->p7 }
 {p6-a7 p6-se p7-se SE->p7 }
 {a6-a7 p6-a7 p6-se p7-se }
 {a6-a7 p6-se A6->a7 p7-se }
 {a6-a7 p6-se p7-se A6->p7 }
 {a6-a7 p6-se SE->a7 p7-se }
 {a6-a7 p6-se p7-se SE->p7 }
 {a6-a7 p6-a7 p6-se p7-ed }
 {a6-a7 p6-se A6->a7 p7-ed }
 {a6-a7 p6-se p7-ed A6->p7 }
 {a6-a7 p6-se SE->a7 p7-ed }
 {a6-a7 p6-se p7-ed p7-se }
 {a6-a7 p6-se p7-ed SE->p7 }

The Total Tetrad Residual, or TTR, is: 0.37422

Figure 8-64: Suggested Sets For Skeleton

According to the suggested sets, it appears that a6-a7 is the best single trek to add to the skeleton. This trek can be realized as a direct cause from the value of a6 to the value of a7, or as a correlated error, but it cannot be interpreted as a cause from a7 to a6. We regard the correlated error as the most plausible realization of this suggested trek addition. The second suggested set contains a6-a7 and it also involves a pair of treks from se to

the measures of powerlessness at two different times. It is certainly plausible that a measure of SES and a measure of powerlessness have correlated errors. It is natural to suppose that if SES is correlated with a measure of powerlessness at one time then it is correlated with the same measure at a later time, and the second set of suggested treks agrees with that judgement, which we can think of as a symmetry requirement imposed on our models. The socio-economic index cannot, it would seem, be either a cause or an effect of the *responses* given to questionnaire items, and so it seems less reasonable to implement the suggested treks by means of directed edges connecting p6, p7 and se.

An immediate realization of the second set of suggested treks is therefore a very plausible elaboration of the initial model, namely:

$$a6 \text{ C } a7, \quad p6 \text{ C } se, \quad p7 \text{ C } se$$

This model has a chi square probability of .9103. In view of the goodness of fit, and the fact that the models in the third unit of suggested sets of trek additions will all fail to explain one or more tetrad equations that hold at a significance level of .49, we do not consider models from the third unit. In fact, it also turns out that all of the plausible models from the third unit of suggested trek additions have chi square probabilities lower than .9103.

8.9.3. Searching with the Rtrr Chart

The significance values for the first two suggested sets are .13 and .49 respectively. We compromise and choose .4 as the significance level for the Rtrr chart. The chart produces a number of good candidate additions from which we form the first level of a search tree. The fragment of the chart containing *only the best additions* is illustrated below:

The graph analyzed in this example is:

A6→a6 A6→p6 A6→A7 A7→a7 A7→p7 SE→ed SE→se SE→A6 SE→A7

The significance level is: 0.4000

The sample size is: 932

Edge	Rttr	D(I-H)	I(H-I)
a6→ a7	0.276	6	1
a6 C a7	0.249	5	0
p6→ a7	0.276	6	1
p6→ p7	0.229	5	2

The Total Tetrad Residual, or TTR, is: 0.37422

I-H, the total number of equations implied by the model that do not hold in the data, is: 11

H-I, the total number of equations that hold in the data but are not implied by the model, is: 1

The number of equations that hold in the data is 5

The number of equations explained by the model is 4

Figure 8-65: Reduced Rttr Chart For The Skeleton

What makes these additions to the initial model superior to alternative elaborations is that they have small values for I(H-I) and relatively large values for Rttr. The small values for I(H-I) mean that most of the equations that hold at this significance level and are implied by the initial model are also implied by the revised model. So these modifications are best by Spearman's Principle. The large values of Rttr mean that these models have smaller TTR values than do other one step modifications of the initial model, so they are best by Thurstone's Principle.

We now request an Rttr chart for each of the models produced by adding one of the edges above to the initial model. These Rttr charts give us information about the second stage or level of modification of the initial model. After locating the superior candidates in each Rttr chart, we form the second level of the Rttr search tree. We show this in Figure 8-66 below.

Root -----	Level 1 -----			Level 2 -----		
	Edge -----	TTR ---	Pi --			
				ed->a6	.007	.817
				a6->se	.007	.817
				se->a6	.007	.817
	a6->a7	.098	.737	a6 C se	.040	.743
				p6->se	.007	.817
				p6 C se	.038	.774
				a7->se	.007	.817
				a6->se	.014	
	a6 C a7			a7->se	.014	-
	(T->a6,T->a7)			T->se	.022	-
Initial Model				p6 C se	.053	-
				a6->se	.007	-
	p6->a7			p6->se	.007	-
				a7->se	.007	-
				se->a6	.033	
	p6->p7			ed->p6	.001	
				se->p6	.001	
				p7->ed	.021	

Figure 8-66: Two Levels of the Rtrr Tree

Proceeding to a third level in this way would require running TETRAD on the 18 second level models and finding their Rtrr charts. This is perfectly feasible, and would require perhaps two hours. It could be done easily in batch mode. An alternative procedure, however, is to reduce the number of level 2 alternatives by computing their chi square statistics, keeping only the those second stage models with significant p values, and then proceeding with the TETRAD analysis. Using EQS on a personal computer (without a coprocessor) to estimate and test the eighteen models requires several hours. To illustrate the possibility, however, we ran EQS on all eighteen models in batch mode. The computer did the work overnight while we slept. The results are given below.

Edges Added To Skeleton	DF	p(X ²)
1) a6→a7, ed→a6	4	0.0106
2) a6→a7, a6→se	4	0.0436
3) a6→a7, se→a6	4	0.0733
4) a6→a7, se C a6	4	0.0665
5) a6→a7, p6→se	4	0.0293
6) a6→a7, p6 C se	4	0.0565
7) a6→a7, a7→se	4	0.0492
8) a6 C a7, a6→se	4	0.5097
9) a6 C a7, a7→se	4	0.4057
10) a6 C a7, a6 C se, a7 C se	3	0.8017
11) a6 C a7, p6 C se	4	0.6539
12) p6→a7, a6→se	4	<.001
13) p6→a7, p6→se	4	<.001
14) p6→a7, a7→se	4	<.001
15) p6→p7, se→a6	4	<.001
16) p6→p7, ed→p6	4	<.001
17) p6→p7, se→p6	4	<.001
18) p6→p7, p7→ed	4	<.001

Figure 8-67: Intermediate Results For Level Two

The tree can be pruned by eliminating the last two nodes on level 1 and all of their descendants. That leaves us with the smaller tree:

Root	Level 1			Level 2		
----	Edge	TTR	Pi			
	----	---	--			
Initial Model				ed→a6	.007 .817	
				a6→se	.007 .817	
				se→a6	.007 .817	
		a6→a7	.098	.737	a6 C se	.040 .743
				p6→se	.007 .817	
				p6 C se	.038 .774	
				a7→se	.007 .817	
				a6→se	.014 -	
		a6 C a7			a7→se	.014 -
		(T→a6, T→a7)			T→se	.022 -
				p6 C se	.053 -	

Figure 8-68: Pruned Tree

TETRAD indicates that in most of these cases searching for a third edge would be of marginal value. Look at the nodes on level 2 under $a6 \rightarrow a7$. Most have a TTR of .007. That is because they imply only a single tetrad constraint, which holds very, very closely. Consider, for example, Figure 8-69, which contains the statistical information for the first graph on level 2, namely:

Initial Model + $a6 \rightarrow a7$ + $ed \rightarrow a6$

The figure contains only information on those equations that hold or that are implied. Only one equation is implied by the model, and the tetrad difference in the sample is only .0075.

The graph analyzed in this example is:

$a6 \rightarrow a7$ $ed \rightarrow a6$ $A6 \rightarrow a6$ $A6 \rightarrow p6$ $A6 \rightarrow A7$ $A7 \rightarrow a7$ $A7 \rightarrow p7$ $SE \rightarrow ed$ $SE \rightarrow se$ $SE \rightarrow A6$ $SE \rightarrow A7$

The significance level is: 0.4000

Tetrad Equation	Residual	Impld.	Held	P(diff.)
$a6 p6, a7 ed = a6 a7, p6 ed$	0.0014		y	0.9436
$a6 p7, ed p6 = a6 ed, p7 p6$	0.0068		y	0.6085
$a6 ed, se a7 = a6 se, ed a7$	0.0007		y	0.9399
$p6 a7, p7 ed = p6 p7, a7 ed$	0.0081		y	0.5296
$p6 ed, se p7 = p6 se, ed p7$	0.0075	y	y	0.4903

Figure 8-69: Statistical Information For Skeleton + $a6 \rightarrow a7, ed \rightarrow a6$

The implied tetrad equation holds closely. We prefer models that *explain* the equation to models that do not imply it. There is, therefore, no reason to add any further directed edge or correlated error to this modification of the initial model.

The situation is nearly the same for the other nodes. The node under $a6 \rightarrow a7$ that adds $a6 C se$ implies only two equations, each of which have very high p values.

Only the node that adds $p6 C se$ to $a6 \rightarrow a7$, the last node in Figure 8-68, seems worthy

of further modification. That model implies only two tetrad equations, each of which has a p value less than .4. Remarkably, there are three additions to this graph which defeat the implication of these and only these two equations, and one of them is p7 C se. We therefore arrive again at one of the models obtained from the Suggested Sets of trek additions.

Turning to the nodes on level 2 under a6 C a7, which did much better on a chi square test than did those under a6→a7, we need not search under a6→se or a7→se. The graphs corresponding to those nodes imply only two equations, both of which hold at .49 or better. The other two nodes still deserve further consideration.

The third node under a6 C a7, called T→se should be explained. TETRAD will only accept directed graphs, not graphs with undirected lines representing correlated errors. *To give TETRAD a model with a correlated error, the correlated error must be represented as a latent common cause.* If we then ask for an Rtrr chart, TETRAD assumes T is a latent variable just like A6, A7 and SE. It calculates the effect of adding extra edges from T to other indicators and includes them in the Rtrr chart. T→se is just such an edge. The graph corresponding to this node actually adds three free parameters to the initial model. The following two models are both realizations of this node.

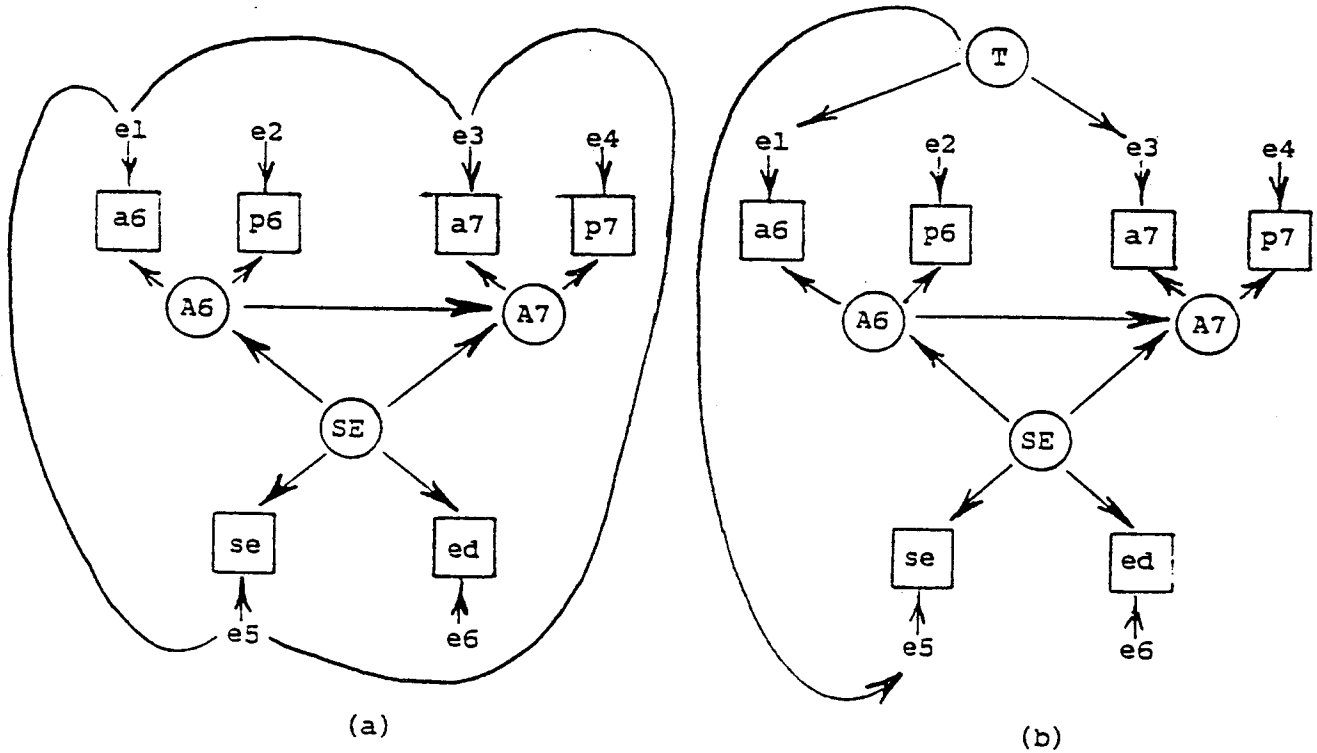


Figure 8-70: Two Models That Realize Node 3 Under a6 C a7

The model in Figure 8-70-b is given to TETRAD, requesting standard output, with the following results:

The graph analyzed in this example is:

```
SE->ed SE->se SE->A6 SE->A7 A6->a6 A6->p6 A6->A7 A7->a7 A7->p7  
T ->a6 T ->a7 T ->se
```

The Total Tetrad Residual, or TTR, is: 0.02236

I-H, the total number of equations implied by the model that do not hold in the data, is: 0

H-I, the total number of equations that hold in the data but are not implied by the model, is: 2

The number of equations that hold in the data is 5

The number of equations explained by the model is 3

Figure 8-71:

There is no need to search for a further addition to this model. It explains three of five equations that hold at .4 or better, and it implies no equation that does not hold at .4 or better.

The final node under a6 C a7, p6 C se, is represented for TETRAD as the graph in the output file we show in Figure 8-72. The Rtrr chart for this node is unambiguous. There is a single best addition, which we show below.

The graph analyzed in this example is:

SE->ed SE->se SE->A6 SE->A7 A6->a6 A6->p6 A6->A7 A7->a7 A7->p7
T ->a6 T ->a7 T1->p6 T1->se

Edge	Rttr	D(I-H)	I(H-I)
p7 C se	0.038	3	0

The Total Tetrad Residual, or TTR, is: 0.05387

I-H, the total number of equations implied by the model that do not hold in the data, is: 3

H-I, the total number of equations that hold in the data but are not implied by the model, is: 2

The number of equations that hold in the data is 5

The number of equations explained by the model is 3

Figure 8-72: Rttr Chart For Final Node Under a6 C a7

If we add p7 C se, the resulting model will explain three of five equations that hold at .4, it will imply no equation that holds at .4, and it will have a TTR of .0158, which is very low. It is striking that the program and heuristics use structural criteria to find the addition that is also preferable on grounds on symmetry.

8.9.4. Goodness of Fit Results

Using EQS in batch mode, p values for the chi square statistic of all of the final models described above were obtained. The results are given below:

Edges Added To Skeleton	TTR	Equations Explained (at .4)	p(X ²)
None	.3742	4 of 5	<.001
From Suggested Sets			
1) a6 C a7, se C p6, se C p7	.0158	3	0.9103
From the Rtrr Chart			
1) a6→a7, ed→a6	.007	1	0.0106
2) a6→a7, a6→se	.007	1	0.0436
3) a6→a7, se→a6	.007	1	0.0733
4) a6→a7, se C a6	.014	2	0.0665
5) a6→a7, p6→se	.007	1	0.0293
6) a6→a7, p6 C se	.027	2	0.0565
6a) a6→a7, p6 C se, p7 C se	.006	2	0.0895
7) a6→a7, a7→se	.007	1	0.0492
8) a6 C a7, a6→se	.014	2	0.5097
9) a6 C a7, a7→se	.014	2	0.4057
10) a6 C a7, a6 C se, a7 C se	.022	3	0.8017
11) a6 C a7, p6 C se	.053	3	0.6539
11a) a6 C a7, p6 C se, p7 C se	.016	3	0.9103

Table 8-6: Goodness of Fit Results: All Models

Of the models obtained with the Rtrr chart, we can eliminate models 1 through 7 because of their chi square p values. Models 8 and 9 do well on the chi square test, but they violate the symmetry constraint and are not as plausible as the other promising models. Model 10 is a good model. It satisfies the symmetry constraint, it is plausible, and it does extremely well on the chi square test. Model 11a is the same model as model 1 from the suggested sets.

These considerations leave us with two models that seem the best on all criteria.

Edges Added To Skeleton	TTR	Equations Explained (at .4)	p(X ²)
None	.3742	4 of 5	<.001
1) a6 C a7, se C p6, se C p7	.0158	3	0.9103
2) a6 C a7, a6 C se, a7 C se	.0226	3	0.8017

Figure 8-73: Most Promising Models

While Model 1 has a nice symmetry between the indicator *se* and the indicator *powerlessness* measured at two different times, it asserts that there is no correlation between the error terms for *powerlessness* at the different times. It would seem more plausible to add another correlated error between *p6* and *p7*. If we do, we get this result

Edges Added To Skeleton	TTR	Equations Explained (at .4)	p(X ²)
1a) a6 C a7, se C p6, se C p7, p6 C p7	.0008	1	0.9936

8.9.5. Conclusion

With TETRAD's help we have found three models which perform substantially better on a chi square test than the one standardly exhibited in causal modelling texts. Two of these models lose a single degree of freedom in comparison with the model suggested in the LISREL and EQS manuals.

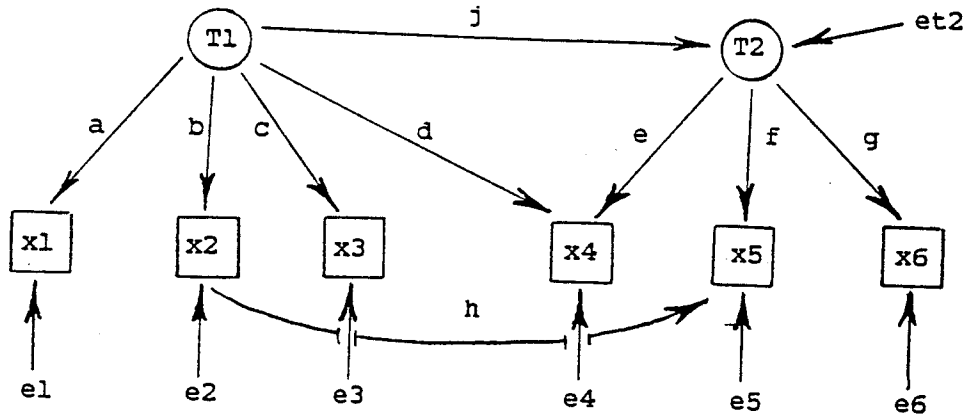
In fact, however, we have done something else at least equally important: *We have performed a search for alternative models.* In order not to make the discussion even more lengthy, we have not described all of the details of that search. For example, we have ignored all of the suggested sets of treks the program produced on the third iteration of the Suggested Sets procedure. We did so because the model found at the second iteration will explain more equations that hold at a high significance level (.4) and because it has an extraordinarily good fit. In fact, all of the models obtained from consideration of the third unit of suggested sets have lower chi square p values than does the model obtained from the second unit.

This case also exhibits something of the role that background knowledge can play in causal modelling with TETRAD. At several points in the analysis we appealed to the knowledge that *a6* and *a7*, and similarly *p6* and *p7*, were the same indicator measured at different times. We also appealed to our knowledge of the time order in eliminating possible causal connections.

8.10. A SIMULATED CASE

8.10.1. Introduction

Causal modelling takes place amidst a host of assumptions. The investigator must assume the relationships among variables are essentially linear, he or she must assume that each variable explicitly modelled has an independent source of variation, etc. Despite tests for various assumptions, with real empirical data we cannot be certain that the assumptions are met. One important kind of test of a discovery procedure is how it fares with such data, and how the results obtained compare with results obtained by researchers who use other procedures. But another important test of a discovery procedure concerns its performance when the assumptions of linear modelling are known to be satisfied. We consider a case in which the data are generated artificially, and the various linearity and stochastic assumptions, as well as the initial model, are known to be correct. We used SYSTAT BASIC to generate data from the statistical model shown in Figure 8-74, with $n = 5,000$. The equations expressing the causal dependencies can be read from the graph in the usual way. We generated values for exogenous variables with a pseudo-random number generator. The data and a fragment of the model (two edges were omitted) are then submitted to a TETRAD analysis. The results reported here are not really a test, since they were not done blind, but the results are sufficiently unambiguous that we are confident that they could be repeated in similar studies in which the TETRAD user did not know the missing edges beforehand.



where:

T1, e1-e6, and et2 are distributed $N(0,1)$.

a = .865; b = .458; c = 1.217; d = .398;
 e = .567; f = .854; g = .772; h = .593; j = .626;

Figure 8-74: Model That Generated The Data

The correlations generated from this model, and a skeleton that does not include the edges T2→x4 and x2→x5 were given as problem 3 in Chapter 6.

8.10.2. The TETRAD Analysis

Recall that the skeleton in problem 3 from Chapter 6 is as follows.

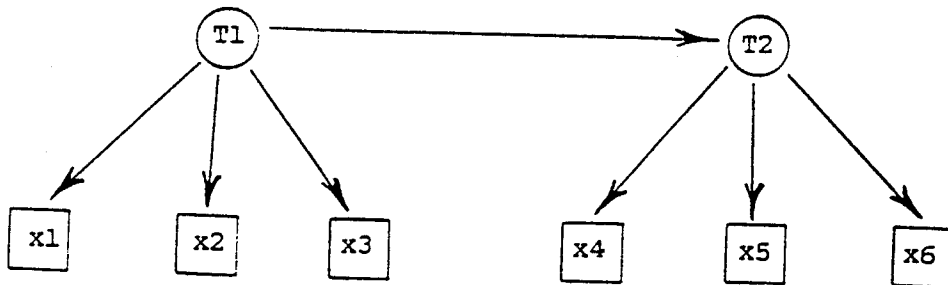


Figure 8-75: Skeleton

Giving this skeleton to TETRAD along with the correlations, we ask for sets of suggested edges. The first unit of suggested sets contains enough sets, and each of them large enough, that we need not worry about a second unit.

The covariance file used for this run is named: sim.dat
 The graph file used for this run is named: sim.g
 The name of this output file is: sim.ol

The graph analyzed in this example is:
 T1->x1 T1->x2 T1->x3 T1->T2 T2->x4 T2->x5 T2->x6

The significance level is: 0.0500

The sample size is: 5000

Sets of suggested treks at significance level = 0.0000

```
{x2-x5 x4-x5 x4-x6 }
{x2-x5 x4-x5 x5-x6 }
{x2-x5 x4-x6 x5-x6 }
{x2-x5 x4-x5 T1->x4 }
{x2-x5 x4-x6 T1->x4 }
{x2-x5 x4-x5 T1->x5 }
{x2-x5 x5-x6 T1->x5 }
{x2-x5 x4-x6 T1->x6 }
{x2-x5 x5-x6 T1->x6 }
{x2-x5 T1->x4 T1->x5 }
{x2-x5 T1->x4 T1->x6 }
{x2-x5 T1->x5 T1->x6 }
```

The Total Tetrad Residual, or TTR, is: 1.37366

I-H, the total number of equations implied by the model that do not hold in the data, is: 15

H-I, the total number of equations that hold in the data but are not implied by the model, is: 2

The number of equations that hold in the data is 14

The number of equations explained by the model is 12

Figure 8-76: Suggested Sets For the Skeleton

The sets all contain the trek x2-x5, so we expect it to be in the original model and to do well in an Rtrr chart. Treks in these sets are suggested at a significance level of .000, which means that all tetrad equations implied by the initial model and having a p value greater than zero are also implied by any model that adds a suggested set of treks (and

only those treks).

We now ask for an Rtrr chart with the default significance level of .05.

The graph analyzed in this example is:

T1->x1 T1->x2 T1->x3 T1->T2 T2->x4 T2->x5 T2->x6

The significance level is: 0.0500

Edge	Rtrr	D(I-H)	I(H-I)	Pi
x2-> x5	1.182	12	0	0.753
x5-> x2	1.088	8	4	0.693
x2 C x5	1.062	8	0	0.676
x4-> x5	0.383	6	0	0.244
x5-> x4	0.383	6	0	0.244
x4 C x5	0.383	6	0	0.244
x4-> x6	0.420	6	0	0.268
x6-> x4	0.420	6	0	0.268
x4 C x6	0.420	6	0	0.268
T1-> x4	0.344	6	0	0.219
x5-> x6	0.344	6	0	0.219
x6-> x5	0.344	6	0	0.219
x5 C x6	0.344	6	0	0.219
T1-> x5	0.420	6	0	0.268
T1-> x6	0.383	6	0	0.244

Figure 8-77: Rtrr Chart For the Skeleton at .05

x2->x5 is by far the best edge. We proceed to add it and request another Rtrr chart. After eliminating inferior candidates, we arrive at 6 possible edges. The results for these 6 are as follows.

The graph analyzed in this example is:

x2->x5 T1->x1 T1->x2 T1->x3 T1->T2 T2->x4 T2->x5 T2->x6

Edge	Rttr	D(I-H)	I(H-I)	Pi
x4-> x5	0.154	3	0	0.729
x4 C x5	0.154	3	0	0.729
T1-> x4	0.154	3	0	0.729
x6-> x5	0.154	3	0	0.729
x5 C x6	0.154	3	0	0.729
T1-> x6	0.154	3	0	0.729

The Total Tetrad Residual, or TTR, is: 0.19182

The Pi value for the model is: 0.75285

I-H, the total number of equations implied by the model that do not hold in the data, is: 3

H-I, the total number of equations that hold in the data but are not implied by the model, is: 2

The number of equations that hold in the data is 14

The number of equations explained by the model is 12

Figure 8-78: Rttr Chart For the Skeleton + x2->x5

Notice that the Pi value for the model that adds x2->x5 is .753. Any further addition will reduce the Pi value, so searching for a third addition may be fruitless, a conclusion which agrees with the fact that the data were generated from a model containing two and only two additional edges.

8.10.3. Goodness of Fit Results

Our best candidate models are the six from the Rttr chart above. Running these models through EQS, we get the following results.

Edge Additions To Skeleton	TTR	Number of Equations Exp.	$p(X^2)$
Skeleton	0.229		<.001
1) x2->x5, x4->x5	0.038	12	<.001
2) x2->x5, x4 C x5	0.038	12	<.001
3) x2->x5, T1->x4	0.038	12	0.842
4) x2->x5, x6->x5	0.038	12	0.842
5) x2->x5, x6 C x5	0.038	12	0.842
6) x2->x5, T1->x6	0.038	12	<.001

Figure 8-79:

Only three of these models have acceptable chi square values, and they include the correct model.

8.10.4. Conclusion

This example, unlike the examples in Chapter 7, do not constitute a real test of the TETRAD program, since we knew the correct answer. We are confident that an experienced user would do equally well blind with similar problems. When multiple edges are missing the TETRAD program will, if the initial model is correct and the sample sizes are large, often reduce the number of alternative models to a handful, including the correct model, but it will not usually produce the correct model uniquely.



9. A BRIEF HISTORY OF HEURISTIC SEARCH IN APPLIED STATISTICS

The methods that are used in the TETRAD program derive from work in statistics, biometrics and psychometrics that is now long forgotten, but once captured the attention of some of the best workers in these subjects. That work forms a fascinating piece of the history of applied statistics. The story that follows is certainly not complete, especially in its recounting of more recent work.

We may forget how closely the development of statistics and probability was sometimes tied to the desire to be able to infer causes from frequencies. George Boole, for example, seems to have thought that the principal application of his algebra was to enable one to infer causes from probabilities, and to that end he embedded his algebra in a rather idiosyncratic theory of probability. His work was roundly, and from a modern point of view rightly, criticized by Cambridge Bayesians. Toward the close of the 19th century, Galton and Pearson developed the ideas of regression and correlation to attempt to describe the statistical manifestations of genetic inheritance. Pearson remained a major force in statistics after the turn of the century, and his students, such as Yule, made major contributions to the subject.

Charles Spearman was an English psychologist who sought to combine the psychometric innovations of workers such as Binet with the new statistical methods that were emerging from mathematicians such as Pearson. Pearson did not like the result, and he and Spearman remained antagonists for a long while. In 1904 Spearman published two papers. One, "The Proof and Measurement of Association Between Two Things" introduced a "correction for attenuation" in the estimation of the correlation coefficient between measures that are severally subject to error. Pearson thought Spearman's mathematics hopelessly obscure and wrote that "Perhaps the best thing at present would be for Mr. Spearman to write a paper giving algebraic proofs of all the formulas he has used, and if he did not discover their erroneous character in the process, he would at least provide tangible material for definite criticism, which it is difficult to apply to mere unproven assertions."

The other paper Spearman published, [101], "General Intelligence Objectively Determined and Measured", drew less notice but was more important. In it, from a variety of tests with a small sample of London school boys, Spearman argued that there is one common component to all manifestations or tests of "intelligence". Reading Spearman's paper nowadays one is struck by the bizarre choice of tests, many of which involved motor skills or sensory discrimination, and the absurd sample sizes, but that is not the point. The

point is Spearman's *argument*, for it was the argument that caught the imagination of statisticians and psychometricians.

Spearman's argument, given rather obscurely in this first paper, but more clearly in subsequent papers, was as follows:

Suppose there are four measured variables x_1 , x_2 , x_3 and x_4 , and their correlations (Pearson product-moment correlations, that is) have values that satisfy three equations:

$$\begin{aligned}\rho_{12}\rho_{34} - \rho_{13}\rho_{24} &= 0 \\ \rho_{12}\rho_{34} - \rho_{14}\rho_{23} &= 0 \\ \rho_{13}\rho_{24} - \rho_{14}\rho_{23} &= 0\end{aligned}$$

These constraints on the correlations can be explained in a powerful way by supposing that each of the measured variables is affected by a single common unmeasured variable, and that each of the measured variables is also affected by an unmeasured variable specific to it and having no effect on the other measured variables. The algebra is easy:

$$\begin{aligned}x_1 &= a_1G + s_1 \\ x_2 &= a_2G + s_2 \\ x_3 &= a_3G + s_3 \\ x_4 &= a_4G + s_4\end{aligned}$$

G , the s_i and the x_i are random variables, the a_i are undetermined constants. Multiplying each side of the first equation by the corresponding side of the second equation and averaging we get:

$$\text{Exp}(x_1x_2) = \text{Exp}(a_1a_2G^2 + a_1s_2G + a_2s_1G + s_1s_2)$$

For algebraic simplicity assume that all of the variables have mean 0. Since there is assumed to be no causal connection between the several s_i variables, or between these variables and G , their covariances should vanish, leaving

$$\text{Cov}(x_1x_2) = a_1a_2\text{Var}(G)$$

Analogous equations are obtained in the same way for the other covariances of the measured variables. Hence, for example:

$$\text{Cov}(x_1x_2)\text{Cov}(x_3x_4) - \text{Cov}(x_1x_3)\text{Cov}(x_2x_4) = a_1a_2a_3a_4\text{Var}(G)^2 - a_1a_3a_2a_4\text{Var}(G)^2 = 0$$

The correlation coefficient is just the covariance divided by the product of the standard deviations of the correlated variables, so the difference of the products of the correlations coefficients must also vanish.

Spearman's point was that the assumption of one common factor would explain why these constraints on the correlations, or *tetrad equations* as they came to be called, were satisfied in the population. The thing that must be emphasized, although it is simple enough, is this:

The one factor model implies that the three tetrad equations will hold in the population no matter what the values of the coefficients a_i may be.

The converse is also true, Spearman thought. That is to say, if correlations among four variables satisfy the three associated tetrad equations, then the correlations can be generated from some one factor model, with appropriate values for the linear coefficients. The three tetrad equations are the necessary and sufficient statistical conditions for the admissibility of a one factor explanation of the correlations. This amounts to saying that a one factor model implies no other independent constraints on the population covariance matrix besides the vanishing of all tetrad differences.

Analogous arguments apply for any larger number of measured variables. If there are n measured variables then the number of possible tetrad equations among these variables is 3 times the number of ways of choosing 4 from n . If the correlations are due to the action of one common cause, then all of the tetrad equations must hold, regardless of the values of the linear coefficients and, conversely, if all of the tetrad equations hold, then the correlations can be generated by a one factor model with appropriate values of the linear coefficients.

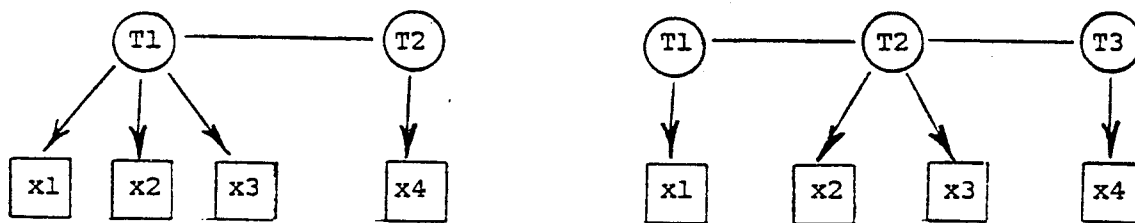
Spearman realized that the three tetrad equations could be derived from alternative models with multiple latent factors, provided the linear coefficients were chosen to have appropriate values, but only the one factor model, he seems to have thought, would imply the tetrad equations no matter what the values of the coefficients might be. Spearman's dominant methodological idea, never fully articulated, seems to have been that the best

explanation is one which generates constraints found in the population measures without having to assume special values for its parameters. In other terms, he preferred those explanations which were robust or resilient over changes in the specification of the value of free parameters.

Whether Spearman realized it or not, he was appealing to a methodological idea that runs throughout the sciences. Kepler gave an analogous argument for the superiority of Copernican to Ptolemaic astronomy [31]. In the same years that Spearman was applying the argument to psychometrics, Harold Jeffreys, who was later to make distinguished contributions to statistics, and Arthur Eddington were making an analogous argument for Einstein's general theory of relativity.

In the years after 1904 Spearman and his students published a number of studies all of which supported his hypothesis of "General Intelligence", and critics such as Thorndike published contrary studies. Some critics, notably G. Thomson argued that the tetrad equations need not be explained by a single common factor, but could instead be satisfied if the test items were produced by a sufficiently large number of completely independent factors. The argument was rather muddled.

Neither Spearman nor his students nor his critics seem to have noted that there are a number of alternative models that have the same relationship to the tetrad equations as does the one factor model. Pictorially, both of the following models entail that three tetrad equations will hold in the population, no matter what the values of the linear coefficients.



It is hard to know what Spearman would have said about these alternatives. He might have dismissed them on the grounds that they are less simple than the one factor theory, which indeed they are. More likely, he would have held them to be "subtheories" of the one factor model--that is, they amount to saying that the single factor consists of two or three causally connected components. Spearman had nothing against substructure.

In fact, however, the case is even worse, for three tetrad equations among four latent variables are robustly implied by models that have no latent factors, but simply postulate directed effects of the measured variables on one another. For example:

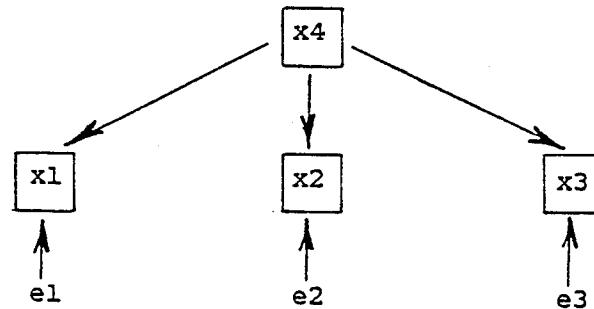


Figure 9-1:

The example is but an instance of Sewall Wright's "path models." Wright's work [Wright], was published in 1921 and drew immediate attention. So far as we know, however, none of the psychometricians of the day, Spearman included, entertained the thought that the responses their subjects gave to earlier items on batteries of mental tests might have an effect on the responses given to later items.

The increasing appreciation of the importance of sampling error made it urgent for Spearman to find a means to determine whether or not a tetrad difference is small enough that it can be plausibly thought to be due to sampling variation from a population in which the true value of the tetrad difference is zero. What was required was a knowledge of the sampling distribution of tetrad differences under the assumption that the underlying distribution of the variables is multinormal. In 1926 Spearman and Karl Holzinger, [Holzinger26], produced an approximate formula for the variance of the sampling distribution of the tetrad residual, or tetrad difference, under the assumption that the difference is zero in the population. A year later they obtained an analogous formula for a collection of tetrad differences, again under the assumption that they are all zero.

Truman Kelley, [51], a psychologist at Stanford and a capable statistician, had earlier derived an approximate formula for the sampling distribution of the ratio of the correlation products rather than the difference of the products. Spearman seems never to have used this statistic, and Kelley himself eventually gave up using it in favor of the tetrad difference. In retrospect, Kelley's strategy seems the better. Spearman was in reality making two inferences. One, that the tetrad residual found in a sample was due to chance and the population tetrad difference was actually zero. Two, that the zero tetrad difference

in the population was due to the causal structure generating the data, and not dependent on the particular values of the linear coefficients. But the tetrad difference is the difference of two small numbers, and the difference of two small numbers is small. Hence, whatever the model, if the linear coefficients were such as to produce small correlations, small tetrad differences would be expected in the population and in the sample. The ratio of two small numbers, by contrast, need not be small. Put in other terms, but still informally, Spearman's statistic has less power to discriminate between small tetrad differences due to causal structure and small tetrad differences due to small correlations.

In 1927 Spearman published the most detailed defense of his theory, together with the statistical formulae. *The Abilities of Man*, [99], drew a lot of attention, including some unwelcome attention from Spearman's old nemesis, Karl Pearson. Pearson wrote unflattering and anonymous assessments of the book in *Nature*, and then, with Margaret Moul, a critical but substantive contribution in *Biometrika*, [80]. Pearson and Moul addressed the question of the sampling distribution for tetrad differences, assuming a vanishing difference in the population and multinormally distributed variables. They did not obtain an exact sampling distribution (and so far as we know the question is still open) and they did not improve on Spearman and Holzinger's formula for the sampling variance. They did, however, observe that the distribution of tetrad differences is certainly not normal. Pearson roundly criticized Spearman, principally on two counts. One, that (using Spearman and Holzinger's formula for the sampling variance) the tetrad residuals in Spearman's data were sometimes quite large. Two, that Spearman had used very small samples (e.g., 50 subjects) and in his statistical analysis had assumed that the sampling distribution of the tetrad differences is normal. Pearson's point was that the sampling distribution is only asymptotically normal, and the sampling variance formula could therefore only be used to decide the significance of a tetrad residual provided the sample sizes were reasonably large, and Spearman's samples weren't.

Spearman didn't get the point, for he replied [100] that he made no assumption about the normality of the distribution of the tetrad differences. But if so, he could not have made the judgements he did about significance, or converted from sampling variance to probable error in the way that was then customary, a transformation that assumes a normal distribution.

John Wishart, [108], next took up the problem of the sampling distribution of tetrad differences. He did not attempt to find an exact sampling distribution, but he did find an exact formula for the variance of the distribution. The exact formula gives a larger standard deviation than does Spearman and Holzinger's approximate formula, and had the effect of making Spearman's judgements in *The Abilities of Man* look a bit more reasonable.

Spearman was interested in tetrads because he saw in them an argument for his psychological theory. But others saw in his work something much more important, namely a general procedure for finding the best linear causal models to account for statistical data. This ingenious work, pursued in different ways by Maxwell Garnett, [31], Truman Kelley, [55] and Karl Holzinger, [47] came to naught, for reasons we will describe subsequently, but it is the real key to our own heuristic search procedures. To explain their ideas, one statistical remark is necessary. Spearman's tetrad equations amount to a simple constraint on the population covariance matrix. (Nowadays, such constraints are given the slightly misleading name of "overidentifying constraints" because they usually determine more than one independent estimate of some parameter of the theory.) Now there are other possible restrictions one could put on the covariance matrix, and in fact the *a priori* possibilities are endless. Garnett and Kelley in effect posed the following problem:

Find for any linear causal model the constraints on the population covariance matrix necessary and sufficient for the admissibility of the model in the sense that the model will imply the constraints for all values of its linear coefficients, and will not imply any further independent constraints.

This is an elegant problem, but one cannot hope to solve it for all possible cases. If, to the contrary, one could classify all of the possible linear models by the constraints necessary and sufficient for them, then by estimating from the sample covariance matrix which constraints are satisfied by the population covariance matrix, one could hope to find the best explanation for any body of data. Kelley fully realized there were far too many possible cases.

Garnett, and especially Kelley, worked out the constraints for a number of models, and the results are reported in Kelley's oddly titled *Crossroads in the Mind of Man*. Kelley's book contains analyses of the following cases:

1. Three variables, one specific factor (i.e. a latent error term) and one common factor (i.e. a latent common cause). (Two of the variables must be perfectly correlated).
2. Three variables, two specific factors and one common factor. (The product of two of the correlations must equal the third.)
3. Three variables, three specific factors and one common factor. (The absolute value of the product of every pair of correlations must be less than the third and one or three of the correlations must be positive.)
4. Three variables, two common factors, no specific factors. (The multiple correlation coefficients of each variable with the other pair must all equal unity.)

5. Four variables, one common factor, four specific factors. (This is Spearman's case.)
6. Four variables, one common factor, three specific factors. (Kelley gives only a sufficient condition.)
7. Four variables, two common factors, four specific factors. (The condition is a slightly complicated inequality)
8. Five variables, two common factors, five specific factors. (The condition is the vanishing of a linear combination of 12 terms each involving a product of five correlations. Kelley calls such a polynomial a "pentad".)
9. Four variables, four specific factors, a common factor and a factor common to x_1, x_2 or a factor common to x_3, x_4 . (The tetrad differences $(12)(34) - (13)(24)$ and $(12)(34) - (14)(23)$ are equal, and hence of necessity the remaining tetrad difference vanishes.)

It should be noted that the last case contains redundant conditions. If, from a foursome of variables, two tetrad differences are equal then it follows necessarily (no matter what the model that generated the data) that the remaining tetrad difference vanishes, and conversely if a tetrad difference vanishes then necessarily the two remaining tetrad differences must be equal to one another.

Kelley's work was very nice, and in some ways the best done by the psychometricians of the time. As a strategy for locating the best linear causal model, however, it was clearly doomed. There are, as the first chapter of this book emphasizes, just far too many possible cases for a complete classification to be carried out. Each case that generates distinct algebraic constraints requires a distinct sampling statistic, and even the indefatigable Kelley did not obtain sampling variances for some of his cases. Kelley realized as much, and he wrote that "we do not have, and certainly cannot readily obtain, criteria for six, seven or more variables at a time." (p. 101).

On the other hand, the cases Kelley did cover offer a kind of *heuristic* guide in the search for models that will best explain the data. However large the model considered, it can be depicted as a directed graph, and the graph will have subgraphs consisting of only three, or four, or five measured variables, and their connections to each other and to latent variables. According to the structure of the subgraphs, constraints of the kinds that Kelley studied will be implied. Thus one might hope to guide the selection of models by comparing the constraints (of the sort Kelley describes) that hold in the data with the constraints of the same sort implied by a hypothetical model, and altering the model if necessary until what holds and what is implied fit together. This is exactly what Kelley himself did.

Kelley applied his analysis to a data set involving nine measured variables for a sample of 140 seventh graders. He settled for trying to fit the model to the vanishing tetrads found approximately in the data. The pentads were just too troublesome to calculate. Holzinger took over the same strategy for the same data in defence of a model that was closer to Spearman's views. A year later, in his *Statistical Resume of the Spearman Two Factor Theory*, Holzinger repeated the same strategy on various examples. He discussed, for example, the case in which there are four measured variables and only one tetrad equation holds, and pointed out, as Kelley had, that the pattern can be accommodated by supposing there is an additional common factor (or as people would say nowadays, correlated error) between either of two pairs of the measured variables. Which pair had the additional common factor, he noted, could be determined if there was an additional, fifth, measured variable.

Holzinger's *Statistical Resume* was nearly the last gasp of tetrad analyses. Spearman's argument that the best explanation is one that robustly implies the constraints found on the covariance matrix had briefly led to a strategy for scientific discovery. The strategy used the most commonly occurring constraint on a covariance matrix, the tetrad equation. It proposed to compute the tetrad equations implied by a model, compare them with the constraints satisfied by the data, and alter the model if there was a misfit. That strategy failed for computational reasons. The determination of the tetrad equations implied by an arbitrary model, even a model with as few as nine or ten variables, was excessively tedious.

The real death of these analyses of constraints on the covariance matrix came at the hands of Hotelling and Thurstone. In the early 1930s Harold Hotelling developed both principal components analysis and canonical correlation analysis. The two techniques enabled researchers to manipulate their data and define constructs which localized the variance of a set of variables, or the correlation of two families of variables. But more important perhaps was Thurstone's contribution.

Thurstone's *The Vectors of Mind* [104] appeared in the middle of the 1930s, and it effectively ended the analysis of tetrad equations and overidentifying constraints among psychometricians and social scientists for another quarter century. Despite the denunciations of factor analysis one hears nowadays, and which as the reader will discover, we endorse, Thurstone's book was elegant and revolutionary. Written in a lucid and courteous style, *Vectors of Mind* was for its time a mathematical *tour de force*, completely in command of the relevant matrix theory, geometry and statistics. Equally important, however, it had a simple and appealing philosophical theme--*science is nothing more than simplification*--and it combined that theme with an easy algorithm for doing science as Thurstone

conceived it, an algorithm that could be carried out by hand with pencil and paper. The procedure was not tied to any one theory of mental or social phenomena, but to a general conception of science. In Thurstone's view science was nothing more than the erection of "constructs" which describe the data more simply, with fewer "degrees of freedom", and he provided an almost magical procedure by which science could be done, theory discovered, progress made.

Spearman's tetrads were discussed by Thurstone, but as an inferior and inadequate special case. The motivation behind them, and the conception of scientific explanation from which they sprang, was never mentioned. Spearman, a mediocre mathematician, a rude and aggressive controversialist, a man tied to one theory and without an automatic method for discovery, had little chance against Thurstone. Neither did Kelley and Holzinger, who quickly converted from the analysis of constraints on the covariance matrix to the methods of factor analysis.

In almost every way, save that it was an effective and easy procedure for producing theories, Thurstone's factor analysis was inferior to the efforts at constraint analysis pursued by Spearman and Kelley and Holzinger and others. Psychometricians had not entertained the possibility that measured variables might have direct effects on one another, or even on latent variables, nor had they analyzed models in which latent variables had *causal* effects on one another, but nothing in the earlier procedures prohibited such analyses. With factor analysis such relationships were impossible because they were beyond the algorithm, and the philosophical views that went with factor analysis discouraged researchers in psychometrics, and in the many other subjects to which factor analysis quickly spread, from even thinking of the possibility.³² Thurstone's factor analysis did not, in any serious sense, conduct a search for alternative theories of the data. Instead it churned out a representative of a collection of statistically equivalent models. Worst of all, perhaps, Thurstone's methods abandoned the insight into scientific explanation that had lain behind Spearman's procedures and that had, for a brief while, threatened to bring statistical modeling into the spirit of the natural sciences. For Thurstone, constraints exhibited by the sample covariance matrix had to be satisfied, but they were to be satisfied by adjusting the linear coefficients--the factor loadings--in whatever way necessary. Thurstone was the Ptolemy of statistical modeling.

³²We think it is fairly clear why none of the psychometricians, from Spearman to Thurstone, considered models in which the latent variables have causal effects on one another. The psychometricians believed their latent factors were genetically determined traits or faculties or abilities. Since the traits were determined by genetics, they could have no causal effects on one another (and neither could they be causally influenced by the manifestations of these traits in observable behavior). Of course, since the traits might have common genetic causes, it was permissible for them to be correlated, as Thurstone's factor analysis allowed. Just why the psychometricians had this conviction is less clear, but two speculations are possible. The source of the study of correlation was Galton's and Pearson's genetic studies, and so it was natural enough for psychometricians to look to biological models. Second, the assumption that the traits manifested by test scores were genetic fit comfortably with their conviction of the superiority of the white race. Truman Kelley, for example, ended his *Crossroads of the Mind* [55] with a discussion of the implications of psychometrics for eugenics.

Thurstone's triumph, and the reasons for it, are evident in Guilford's *Psychometric Methods*, [40], which appeared in 1936. Guilford gives a detailed and sympathetic treatment of Spearman's views and of tetrad methods, before recommending that they be abandoned in favor of Thurstone's factor analysis, which is described in even greater detail. Guilford's reasons for the preference were succinct:

The use of the tetrad is a laborious procedure except when one is dealing with a small number of tests. The number of tetrads increases enormously as the number of tests increases. With 10 tests there are 630 tetrads and 20 tests there are 14,535. More effective and economical methods are now available; hence tetrads have little more than historical importance.

We have no pretense to surveying the history of statistical modeling from the mid-thirties to the present, and we have already ignored the important developments that had taken place in econometrics. But it is worth saying a few words about what happened to the fundamental idea of Spearman's school, namely that linear causal models should be assessed by their ability to robustly entail constraints on the covariance matrix, no matter what the values of their linear coefficients.

Constraints on the covariance matrix came to be known in the economics literature as "overidentifying constraints". For reasons we do not pretend to understand, many econometricians were much more interested in parameter estimation than in theory testing, and overidentifying constraints made consistent, unbiased estimation more difficult. For that reason some econometricians adopted the exact contrary of the views of the Spearman school, and recommended avoiding models that imposed any constraints on the covariance matrix.

The connection between statistical models and causal notions was reinvigorated by Herbert Simon's work in the 1950s, which attempted to derive the causal claims of a statistical model from the identifiability relations of the variables of the model. The analysis seems to have given some applied statisticians encouragement to think explicitly about the connection between statistical models and the causal conclusions people habitually drew from them. In psychology, Donald Campbell's work on inferring causal order from the correlations of time-lagged variables, and in economics the work of Herman Wold, had similar effects. The effect was perhaps most pronounced in sociology, a subject which came rather late to statistical methods and causal models. The leading practitioners, thinking the issues through afresh, in effect rediscovered many of the ideas of the Spearman school.

The clearest example is Hubert Blalock's *Causal Inferences in Nonexperimental Research*, [7] which appeared in 1961. Blalock reinvented the procedure of examining the

constraints on the covariance matrix robustly implied by a linear causal model. The models he considered were not, however, latent factor models of the kind considered by the psychometricians, but were instead path models with no common unmeasured variables, models in fact of the very kind introduced by Sewall Wright. The constraints on the correlation matrix in these cases are not tetrad or pentad equations or the like, but constraints on partial correlations. Blalock makes no reference to any of work done by the Spearman school.

Spearman's ideas were noted by Herbert Costner and O.D. Duncan, however, who considered the patterns of tetrad equations implied by a number of simple latent variable models in which the latent variables have causal effects on one another and each latent variable affects two or more measured indicator variables. They do not refer to the work of Kelley or Holzinger, but their basic idea is much the same. Costner, in collaboration with Ronald Schoenberg, even proposed a more or less systematic procedure for revising linear models with latent variables by locating submodels having two causally connected latent variables each with two measured indicators. Such submodels imply a single tetrad equation, which is no longer implied if a measured indicator of one latent variable has a common factor or correlation with a measured indicator of the other latent variable. If the appropriate tetrad equation is not found to hold empirically, Costner and Schoenberg recommend modifying the model accordingly. Costner and Schoenberg's procedure is still occasionally cited, but does not seem to have been much used, perhaps because it was confined to locating special sorts of erroneous model specifications and was not really algorithmic. Kelley and Holzinger would have liked the work. The development of computer programs such as LISREL that automatically carry out maximum likelihood estimates of parameters and perform significance tests led to a quite different strategy for revising models. The strategy is quite alien to the tradition of the Spearman school and to Costner and Schoenberg's procedure, for it pays no attention to robustly accounting for constraints satisfied by the sample covariance matrix. Instead, the revision procedure locates the fixed parameter for which a small variation will most increase the value of a fitting function, revises the model by adding the corresponding free parameter, reestimates, and tests to see if the difference in chi square values of the original and modified models is significant. With this procedure an initial model may robustly explain constraints on the covariance matrix but the modified model may very well be unable to robustly explain those constraints. If the procedure is not quite as Ptolemaic as Thurstone's factor analysis, neither is it as Copernican as the constraint analyses of the Spearman school. Perhaps we should think of it as Tychonic.

The early psychometricians knew how to compute the covariances or correlations implied by an estimated model, given the variances of the exogenous variables and the values of

the linear coefficients. With Kelley's results they knew how to determine the tetrad equations and certain other constraints for special cases. So far as we know, however, they had no general method for computing covariances and correlations when the models were, in contemporary jargon, non-recursive, that is, when the directed graph of the causal model contained a cyclic path leading from one variable back to that same variable. No such models occurred in psychometrics, but they did occur elsewhere. The general solution to the problem was provided by an outsider, Samuel Mason. His doctoral thesis in the 1950s was devoted to the effects of loops in circuits, and it provided a general procedure for computing the correlations or covariances of a recursive or non-recursive model, given the variances of the exogenous variables and the values of the linear coefficients. Mason's work was elaborated for social scientists by David Heise in his *Causal Analysis*.

The moral to this story is that Spearman, Kelley, Holzinger and others had a good idea about scientific explanation and about how to search for good explanations of nonexperimental data. The idea was not entirely lost, and persists here and there in recent work in statistical modeling. That it never came to fruition was due to computational problems, not to statistical difficulties. What was lacking was a general algorithm that would permit the rapid calculation of the tetrad equations (or other constraints) robustly implied by *any* linear causal model, and would compare those constraints with the constraints approximately satisfied by the sample covariance matrix. That is a lot, of course, but it is more a problem in algorithm design than in statistical inference. With such a procedure embodied in a computer program, researchers would no longer have to bother with lengthy and tedious analyses of the implications of a very few causal models. The implications of any model would be immediately available. With such a procedure researchers would not have to resort to accepting or rejecting a model entirely, but could instead systematically examine the logical workings of parts of a complex model, and modify the parts as need be. A program of this sort would in fact serve as a research tool for determining the sorts of results that Kelley, Holzinger, Costner and Duncan wrung out by hand in a few simple cases. With it, one could locate *general rules* about what modifications of a model would or would not prevent the robust implication of particular tetrad equations, or other constraints. If such general rules could be found, and themselves incorporated in a computer program, we would have the best result of all: with such a procedure, appropriately designed, researchers could implicitly search through millions of alternatives for those modifications of an initial hypothesis that best explain the constraints satisfied by the sample data. TETRAD is such a program.

10. MATHEMATICAL FOUNDATIONS

10.1. THE ALGORITHM

The TETRAD program has four major components.

1. It calculates the vanishing tetrad differences and vanishing partial correlations implied by any model given to it.
2. It determines the tetrad equations and vanishing partial correlations that pass a statistical test at a significance level set either by the user or automatically.
3. It determines which sets of further treks can be added to the model to prevent the implication of "false" tetrad equations without preventing the implication of "true" tetrad equations implied by the initial causal model.
4. For each edge not already present in a model, and for each correlated error between vertices not connected by an edge, it calculates changes in which vanishing tetrad differences, and which vanishing partial differences are implied.

10.1.1. Computing the Implied Equations

The general rules the program uses to compute the tetrad equations implied by a directed graph are as follows.

If u, v, w, x , are four distinct measured variables in the graph, the graph implies the tetrad equation:

$$\rho_{uv}\rho_{wx} = \rho_{ux}\rho_{vw}$$

if and only if

$$\sum_{t \in T_{uv}} \left(\prod_{e \in t} L(e) \right) \sum_{t \in T_{wx}} \left(\prod_{e \in t} L(e) \right) =$$

$$\sum_{t \in T_{ux}} \left(\prod_{e \in t} L(e) \right) \sum_{t \in T_{vw}} \left(\prod_{e \in t} L(e) \right)$$

where e is an edge, $L(e)$ is the label (coefficient) of an edge, and T_{ij} is the set of open treks between i and j .

In the next section, we prove that this rule correctly calculates the tetrad equations implied by any directed graph that represents a linear statistical model.

The TETRAD program uses a modification of a well known path finding algorithm³³ to compute the tetrad equations a model implies. It first calculates the open paths and the open treks between each pair of variables in the graph. The program then calculates, for each pair u,v of measured variables, the trek sum for that pair. The products of the trek sums are then compared to determine whether or not they constitute an algebraic identity. If they do³⁴, the model implies the corresponding tetrad equation. The program proceeds in this way through all foursomes of measured variables.

Computing the Paths between Two Vertices in a Directed Graph

To compute the paths between any pair of variables TETRAD uses an algorithm for computing the cost, $c(v_i, v_j)$, associated with vertices v_i and v_j in a directed graph.

Suppose we have a directed graph $G = (V, E)$. $V = \{v_1 \dots v_n\}$, is the set of vertices in G , and E the set of edges. We have a labeling function $L: (V \times V) \rightarrow S$, where $(S, +, *, 0, 1)$ is a closed semi-ring, and E is a subset of S . We define the label of a path to be the product, $*$, of the labels of the edges in the path, taken in order. The label of a path of zero length is 1, the identity element. The label of a path which does not exist is 0. We want, for all i and j between 1 and n , $c(v_i, v_j)$, or the sum over all paths from v_i to v_j of the path labels. We calculate $c(v_i, v_j)$ as follows.

³³ [1], pp.195-201

³⁴ If there are no treks between a pair of variables on one side of an equation and no treks between a pair on the other side of the same equation, the graph implies the equation and TETRAD counts it as doing so.

```

Begin
1   For 1 <= i,j <= n and i <> j DO  $C^0_{ij} := L(v_i, v_j)$ ;
2   For k:= 1 to n DO
3       For 1 <= i,j <= n DO
4            $C^k_{ij} := C^{k-1}_{ij} + C^{k-1}_{ik} * C^{k-1}_{kj}$ ;
5   For 1 <= i,j <= n DO  $c(v_i, v_j) := C^n_{ij}$ ;
End

```

The complexity of this algorithm is $O(n^3)$, since line 4 is executed n^3 times while lines 1 and 5 are only executed n^2 times.

Calculating the trek sum between two vertices is then simple.

```

Begin
1   For 1 <= i,j <= n DO  $t(v_i, v_j) := c(v_i, v_j)$ ;
2   For k:= 1 to n AND k <> i,j DO
3        $t(v_i, v_j) := t(v_i, v_j) + c(v_k, v_i) * c(v_k, v_j)$ ;
End

```

10.1.2. Suggesting Modifications

To determine the sets of suggested trek additions to an initial model, the program proceeds through all foursomes of measured variables. For each foursome, it locates an appropriate subgraph of the initial model and determines whether or not to recommend the addition of a trek between any pair of variables in the subgraph. The treks may be produced by edges directly connecting two measured variables, or by a common error term connecting the measured variables, or by a directed edge from a latent variable already in the graph to one of the measured variables.

The program's recommendations do not distinguish between directed edges connecting two measured variables in either direction and the introduction of a new error variable

connected to two measured variables. In the program's recommendations, all of these possibilities appear as a measured variable pair in a set of suggested treks. The program does distinguish, however, between these sorts of modifications and the addition of a directed edge between a latent variable and a measured variable. If the program recommends the addition of a directed edge between a latent and a measured variable, then a pair consisting of the latent and the measured variables occurs in the set of recommendations. For example, if TETRAD received the skeleton in Fig. 10-1 and it recommended the pair $[T1 \rightarrow 4]$, then there is only one way to augment the skeleton, the one we show in Fig. 10-2. We suggest this distinction notationally by listing a suggested trek between two measured variables x, y as $[x - y]$, but a suggested trek between a latent variable $T1$ and a measured variable w as $[T1 \rightarrow w]$.

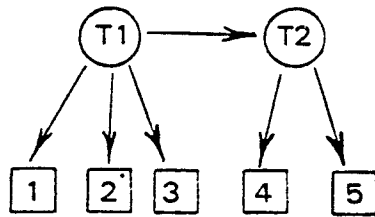


Figure 10-1:

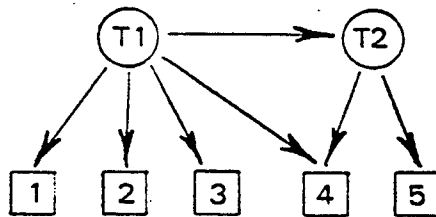


Figure 10-2:

How TETRAD Finds Its Recommendations

The sets of suggestions must strictly satisfy Spearman's principle for a given significance level. TETRAD starts with the significance level set at 0.0 and increases it to a level at which at least one equation is considered false. For that significance level, the program does the following.

TETRAD first cycles through all sets of four measured variables and determines which

equations hold empirically in the foursome at the current significance level. For each foursome, it puts any treks which, if added to the graph, would defeat the implication of equations that hold into a set called DP] (dontpick).

TETRAD then cycles again through all foursomes of measured variables. For each foursome of variables, TETRAD considers whether or not the model implies a tetrad equation in those four variables that does not hold at the given significance level. If it does not, the program goes to the next foursome. If the model does imply an equation that does not hold in the data for those four variables, the basic strategy is to divide and conquer.

For each foursome of variables that implies a tetrad equation that does not hold in the data, the program examines the **subgraph** consisting of

1. The four measured variables, say $x_1 \dots x_4$.
2. The latent variables to which $x_1 \dots x_4$ are adjacent (called parents).
3. The directed edges from parents to adjacent measured variables.
4. All treks, including theoretical variables which are not parents, between the parents.

For example, the graph in Figure 10-3 is skeletal. The subgraphs for variables $x_1-x_2-x_3-x_5$ and variables $x_1-x_2-x_5-x_6$ are shown in Figure 10-4.

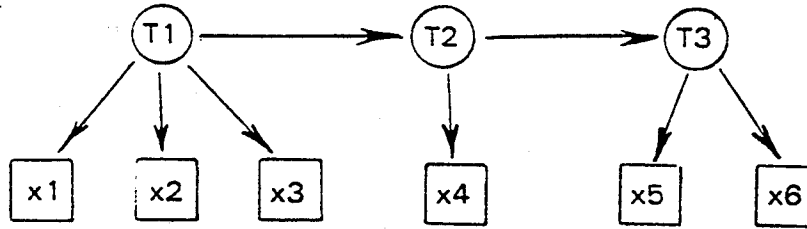


Figure 10-3: Skeleton

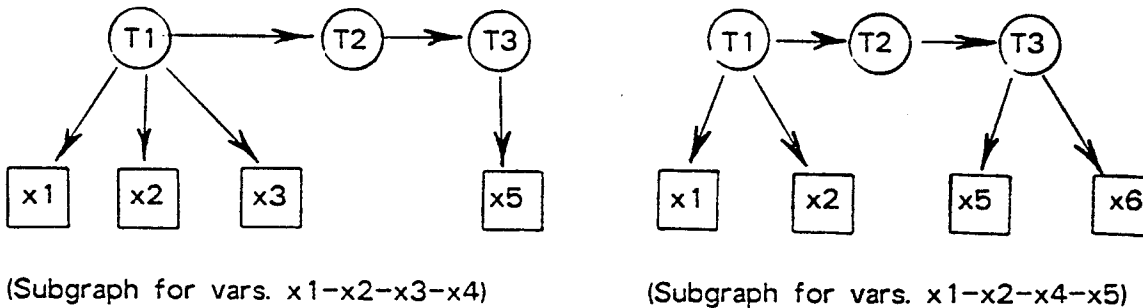


Figure 10-4: Subgraphs

If this strategy is applied to a skeletal graphs, there is a unique parent for each measured variable, and there are no treks between measured variables that do not go through a parent. TETRAD issues a warning that the algorithm may be inaccurate if the graph is not skeletal.

Based on the structure of the subgraph for each foursome, the program assigns the subgraph to one of nine types. The types are really equivalence classes of such subgraphs, where the equivalence relation is restricted to subgraphs that imply at least one tetrad equation.

Two skeletal subgraphs A, B with four measured variables are equivalent if and only if:

1. There is a 1-1 map $F: \{\text{measured and parent variables in A}\} \rightarrow \{\text{measured and parent variables in B}\}$. We consider an edge an ordered pair of vertices, $\langle v_1, v_2 \rangle$. Thus $F\langle v_1, v_2 \rangle = \langle F(v_1), F(v_2) \rangle$. Latent variables in the subgraph are parent variables in this definition only if they are directly connected with one of the four measured variables.
2. x_i is measured in A if and only if $F(x_i)$ is measured in B.
3. T_i is a parent in A if and only if $F(T_i)$ is a parent in B.
4. If T_i is a parent of x_i in A, then $F^{-1}(F(T_i))$ is a parent of $F^{-1}(F(x_i))$ in A.

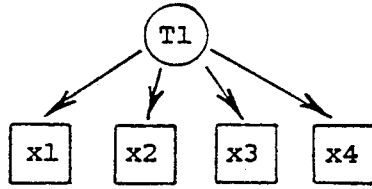
5. A and B imply the same tetrad equations.
6. If any edge(s) from latent to measured variables or from measured to measured variables in A is added to A (forming $A_{\text{augmented}}$), and F(edge(s)) to B, then $A_{\text{augmented}}$ implies the same tetrad equations as $B_{\text{augmented}}$.

For example, all of the subgraphs having two parent variables connected by a trek and having each parent adjacent to two of the four measured variables, form a type.

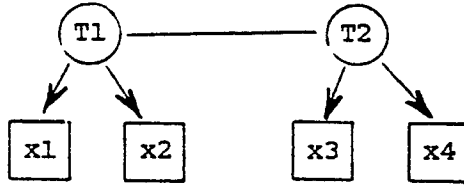
We illustrate the nine types with a representative from each equivalence class. In these illustrations, any number of latent variables may be inserted between the parent variables, so long as the path connections are retained, and no treks are introduced that violate the conditions stipulated in the diagrams.

Directed Edge	=	
Any Path	=	
Any Trek	=	
Latent Parents	=	T1 . . T4
Measured Vars.	=	x1 . . x4

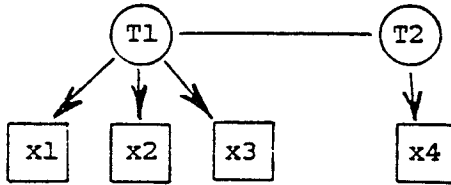
1)
1
Parent



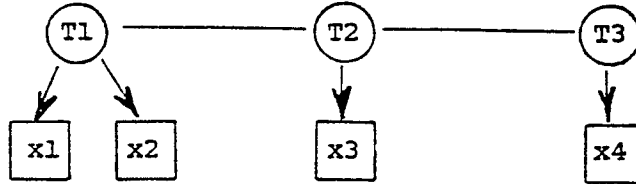
2)
2
Parents



3)
2
Parents

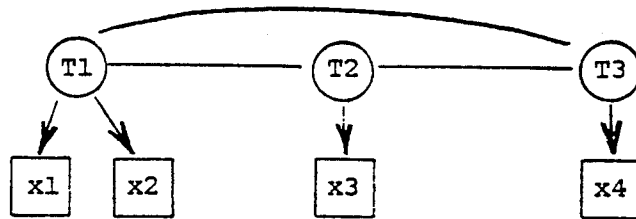


4)
3
Parents



T1 - T3 must be trek connected through T2

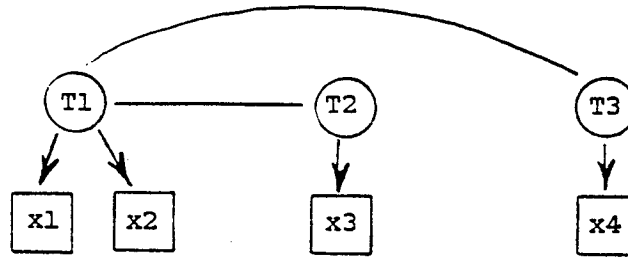
5)
3
Parents



Must be an acyclic graph

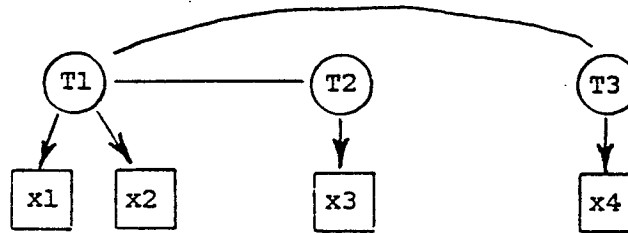
Figure 10-5: Representatives of Subgraph Equivalence Classes

6)
3
Parents



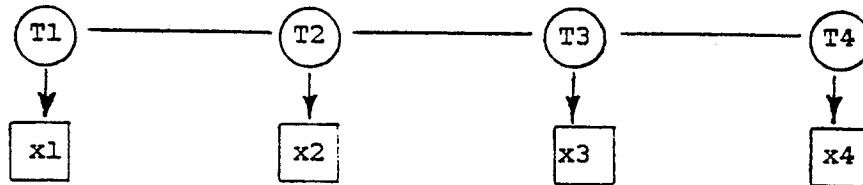
T2 - T3 must be trek connected through T1

7)
3
Parents



T2 - T3 must *not* be trek connected

8)
4
Parents



T1 - T4 must be trek connected through T2 & T3

9)
4
Parents

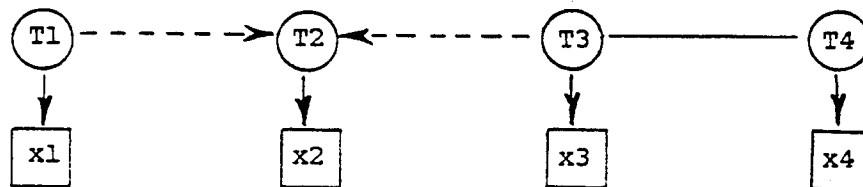


Figure 10-6: Representatives of Subgraph Equivalence Classes (continued)

Based on the type of the subgraph, the program knows all of the additional treks that will defeat the implication of the false equations involving the foursome of measured variables occurring in the subgraph. TETRAD removes any treks from this set that are also in DP, thereby disallowing any candidates for addition that would violate Spearman's principle. After removing the treks that are also in DP, the program forms sets of treks LM_i that are locally minimal in the following sense:

For the subgraph under consideration, each set LM_i , but no proper subset of LM_i , will, when added to the graph given to the program, defeat the implication of as many false tetrad equations involving the measured variables of the subgraph as is possible without defeating the implication of any true tetrad equations implied by the full graph.

As the program proceeds from foursome to foursome it forms globally minimal sets GM_i that have a similar property. For each set GM_i , and no proper subset of GM_i , the set GM_i will defeat as many false equations involving all foursomes considered *so far* as is possible without defeating any true ones. After all foursomes are considered, the sets GM_i are the sets of suggested trek additions.

TETRAD outputs the sets and the first significance level at which the sets of the suggested trek additions are non-empty. It then increases the significance level until the sets of suggested trek additions are different.

10.1.2.1. Computing the Rtrr Values

The Rtrr value for an edge is the amount that the TTR will drop if that edge is added to the graph input. adds that trek to the graph input and recalculates the TTR value. It recalculates the TTR value by determining what paths and treks would be added to a graph if a given edge were added. The difference between the original TTR and the recalculated TTR is the Rtrr for the trek in question. The RTRR value is calculated for each edge that is not in the current model, and for each correlated error between vertices that are not connected by an edge in either direction. Although guaranteed to be accurate, this procedure can be rather time consuming, due to the large number of edges and correlated errors that must be considered. As a result, if you request the Rtrr chart for treks which involve a latent variable, you must expect a substantially longer processing time.

10.2. PROOFS OF CORRECTNESS OF THE ALGORITHMS EMPLOYED BY TETRAD

10.2.1. Introduction

In this section three fundamental results about statistical linear causal theories will be proved. These theorems will be stated precisely later. Informally, they are:

1. Any two theories with the same causal graph imply the same tetrad equations and the same vanishing partial equations.
2. For any theory in which each variable has a unique external source of variance uncorrelated with any other source of variance, the algorithm employed by TETRAD correctly calculates the tetrad equations that the theory implies.
3. For any theory in which each variable has a unique external source of variance uncorrelated with any other source of variance, the algorithm employed by TETRAD correctly calculates the partial equations that the theory implies.

The second point is not obvious for two reasons. When TETRAD determines whether or not a given tetrad equation is implied by a theory, it does not correctly calculate the covariances in that tetrad equation. That is because the covariances are functions of variances of some variables, and TETRAD does not take into account these variances. Correct calculation of the covariances occurring in tetrad equations requires consideration of such variances. Second, in the case of cyclic graphs, the complete calculation would involve taking the sums of products of labels of edges in any trek connecting two variables in a covariance; TETRAD only considers the sums of products of labels of edges in *open* treks (treks without any cycles). The program nonetheless correctly computes the tetrad equations implied by a model.

Before giving the formal definitions and lemmas needed to state and prove these theorems, we will give an example which illustrates them.

10.2.2. Tetrad Equations

Assume that there is a set R , of random variables $\{x_1, x_2, x_3, x_4, x_5, x_6\}$ defined over a sample space Ω with a probability measure P , and a set of five homogeneous equations, each of the form

$$x_i = \sum_{j=1}^6 a_{ij} x_j$$

where each a_{ij} is a constant (possibly equal to zero). The set of non-zero coefficients a_{ij} will be called the **equation coefficients**. For this case

$$x_2 = a_{21}x_1$$

$$x_3 = a_{31}x_1$$

$$x_4 = a_{41}x_1$$

$$x_5 = a_{52}x_2$$

$$x_6 = a_{62}x_2$$

A causal graph D is associated with these equations. The set of random variables R is the set of vertices of the graph. x_i is adjacent to x_j iff x_i causes x_j .

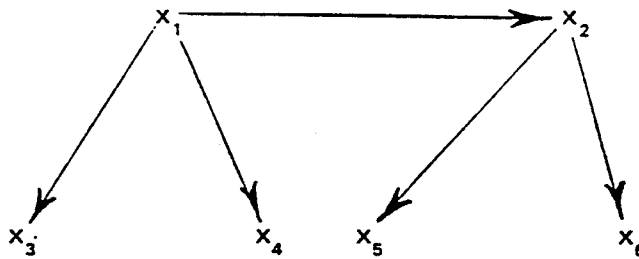


Figure 10-7: Example 1

For each vertex x of positive indegree, x is specified as a function of x_1, \dots, x_n where x_1, \dots, x_n are all vertices adjacent to and into x .

Note that the causal graph completely determines the form of the equations associated with it. However, different sets of equations can be associated with the same graph, since the graph fixes neither the values of the coefficients nor the members of R . Since the members of R are not fixed, neither are the distributions, means or variances of the variables in R .

According to the trek rules (which are described subsequently)

$$\text{cov}(x_5, x_6) = a_{52} a_{62} \text{var}(x_2)$$

Let us call this form of the equation, prescribed by the trek rules, the trek form of the covariance. Each term in this equation is of the form

$$c_i \text{var}(x_i)$$

Let us call c_i a **term coefficient** and an exogenous source of variance an **independent variable**. Note that each term coefficient c_i is a product of powers of the a_{ij} , the equation coefficients. The x_i are not necessarily independent variables (x_2 is not independent).

Since x_2 is not independent, its variance can be written in terms of the variances of variables which are independent. In this case,

$$\text{var}(x_2) = a_{21}^2 \text{var}(x_1)$$

Upon substitution, then,

$$\text{cov}(x_5, x_6) = a_{52} a_{62} a_{21}^2 \text{var}(x_1)$$

When a covariance is written as a sum of variances of independent variables (as above) the covariance will be said to be in **fully expanded form**. Trek form may be, but is not necessarily, the same as fully-expanded form, since some of the variances in the trek form may be variances of non-independent variables. Note that if there is a cycle in the graph, when written in fully-expanded form some covariances will be infinite sums.

In the graph of 10-7

$$\text{cov}(x_5, x_6) = a_{52} a_{62} a_{21}^2 \text{var}(x_1)$$

$$\text{cov}(x_3, x_4) = a_{31} a_{41} \text{var}(x_1)$$

$$\text{cov}(x_3, x_6) = a_{31} a_{21} a_{62} \text{var}(x_1)$$

$$\text{cov}(x_4, x_5) = a_{41} a_{21} a_{52} \text{var}(x_1)$$

Hence,

$$\text{cov}(x_5, x_6) * \text{cov}(x_3, x_4) =$$

$$\text{cov}(x_3, x_6) * \text{cov}(x_4, x_5) =$$

$$a_{31} a_{62} a_{41} a_{52} a_{21}^2 \text{var}(x_1)$$

This equality obviously holds no matter what the values of the equation coefficients are. We shall say therefore that the theory described in Example 1 **strongly implies** the tetrad

equation.

Some of the ideas illustrated in the example will now be defined more formally.

Definition 1: Given an ordered n -tuple $N = \langle c_1, \dots, c_n \rangle$, an object o is in N iff $o = c_i$ for some i between 1 and n inclusive. We shall also write that

$$o \in N$$

This notation is somewhat ambiguous since ϵ is also used to mean set membership, but the context will always make it clear which use of ϵ is intended.

Definition 2: A **digraph** is an ordered pair $\langle V, E \rangle$, where V is a set of **vertices** and E is a set of **edges**. Each edge is either an ordered pair of elements of D or a 1-tuple consisting of a single element of D . If a member of E is an 1-tuple consisting of a single element of D it is called a **degenerate edge**. The first element in a non-degenerate edge is called the **head** and the second element is called the **tail**. In a degenerate edge, the head and tail are both the single vertex occurring in the edge. A non-degenerate edge with a head v_i and a tail v_j is an edge from v_i to v_j ; it is also said that the edge is **out of** v_i and **into** v_j . v_i is **adjacent to** v_j iff there is a non-degenerate edge from v_i to v_j . The **indegree** of a vertex v is equal to the number of distinct non-degenerate edges into v ; the **outdegree** of a vertex v is equal to the number of distinct non-degenerate edges out of v .

Definition 3: Two digraphs $\langle V, E \rangle$ and $\langle V', E' \rangle$ are **isomorphic** iff there exists a 1-1 function Z with domain V such that for each v_i and v_j in V , $\langle v_i, v_j \rangle \in E$ iff $\langle Z(v_i), Z(v_j) \rangle \in E'$.

Definition 4: A **path of length n** is an ordered n -tuple of non-degenerate edges e_1, \dots, e_n where for $1 < i < n$, the tail of e_i is equal to the head of e_{i+1} . As a special case, a **degenerate path** is a 1-tuple consisting of a single degenerate edge. A degenerate path is of length 0. A degenerate edge may appear only in a degenerate path. The head of e_1 is called the **source** of the path; the tail of e_n is called the **sink** of the path. The path is said to **connect** the source to the sink. Two paths P_i and P_j **intersect** iff there exists a vertex in one of the edges of P_i which is also in one of the edges of P_j ; any such common vertex is a **point of intersection**. A **cycle** is a non-degenerate path in which there exists a vertex which is both the head of an edge in the path and the tail of an edge in the path. A path **contains a cycle** iff it has a sub-path which is a cycle. A digraph is **acyclic** if and only if it contains no cycles.

Definition 5: A **trek t** between two vertices v_i and v_j is an ordered pair of paths from a common source, in which the source is the only point of intersection of the two paths. It may be that the source is equal to v_i or v_j , in which case at least one of the paths must be degenerate. t_1 is the first path, and t_2 is the second path. The source of the paths in the trek is called the **source of the trek**.

Definition 6: A set S of equations is **non-redundant** iff there is no subset T of S which implies an equation in S but not in T .

Definition 7: Let a stochastic linear causal theory (SLCT) be $\langle D, (\Omega, P), R, X, L \rangle$ where

1. (Ω, P) is a sample space.
2. R is a set of random variables over (Ω, P) .
3. X is a consistent, non-redundant set of m homogeneous linear equationals in n
4. random variables (x_1, \dots, x_n) where $m < n$ and

$$x_i = \sum_{j=1}^n a_{ij} x_j$$

A non-zero value of a_{ij} is the **equation coefficient** of x_j in the equation for x_i .

5. D is a labelled digraph $\langle R, E \rangle$, (R is the set of random variables which serves as the set of vertices of the graph.) E is a set of directed edges in which there is a non-degenerate edge from x_j to x_i iff $a_{ij} \neq 0$. There is also a degenerate edge from every vertex. There are no non-degenerate edges from a vertex to itself.
6. For each vertex $x \in R$ of positive indegree, X specifies x as a linear function of just x_1, \dots, x_n , where x_1, \dots, x_n are all vertices adjacent and into x . The coefficients of x_1, \dots, x_n are non-zero. If $x \in R$ is of 0 indegree, then there is no equation for x in X .
7. If vertices u and v are not connected by a trek, then v and u are statistically independent.
8. L is a function with domain E such that

$$L(e) = \begin{cases} a_{ij} & \text{if } e \text{ not degenerate, head}(e) = x_i \text{ and tail}(e) = x_j, i \neq j \\ 1 & \text{if } e \text{ is degenerate} \end{cases}$$

Variable x_i is **independent** iff x_i has zero indegree (no edges directed into x_i). Otherwise x_i is **dependent**. (Note that the **property** of independence is completely distinct from the **relation** of statistical independence. The context will make clear in which of these senses the term is used.)

The requirement that the number of equations be less than the number of random variables eliminates the possibility that the value of each random variable is constrained to be a constant. Note that this requirement also implies that there is at least one independent

variable (since otherwise there would be n equations for n variables).

The following lemmas show how to calculate the variance of random variables and covariances between random variables in terms of the variances of independent random variables.³⁵

Lemma 8: If x_1, \dots, x_n are random variables and

$$y_1 = \sum_{j=1}^n a_j x_j$$

and

$$y_2 = \sum_{j=1}^n b_j x_j$$

then

$$\text{cov}(y_1, y_2) = \sum_{i=1}^n \sum_{j=1}^n (a_i b_j) \text{cov}(x_i, x_j)$$

Lemma 9: If x_1, \dots, x_n are random variables and

$$y = \sum_{j=1}^n a_j x_j$$

then

$$\text{var}(y) = \sum_{i=1}^n \sum_{j=1}^n a_i a_j \text{cov}(x_i, x_j)$$

Definition 10: A SLCT is **acceptable** iff in its graph, for every cyclic path P ,

$$\prod_{e \in P} L(e) < 1$$

(We shall adopt the convention that the sum or product over an empty set is equal to 0.)

If a theory is not acceptable, then there is a cycle in which the product of the labels of the edges is greater than 1. This implies that the label of any path that contains an infinite

number of traversals of that cycle will have an infinitely large value. As we shall see later, this in turn would imply that the equations defining some of the random variables diverge, and are not well defined. For this reason, we will restrict our attention only to acceptable theories. Henceforth, when the word "SLCT" is used, it is to be understood that only acceptable SLCTs are being referred to.

Definition 11: A tetrad equation is **strongly implied** by a SLCT $\langle D, (\Omega, P), R, X, L \rangle$ iff it is implied by every theory $\langle D, (\Omega, P), R, X', L' \rangle$ (that is, the theory implies the tetrad equation for every value of the coefficients in X).

Definition 12: The sets of tetrad equations strongly implied by two SLCTs $T = \langle D, (\Omega, P), R, X, L \rangle$ and $T' = \langle D', (\Omega', P'), R', X', L' \rangle$ are **isomorphic** iff there exists a 1-1 function $Q: R \rightarrow R'$ such that a tetrad equation G is strongly implied by T iff the equation obtained by substituting for every variable p in G the corresponding variable $Q(p)$ is strongly implied by T' .

Lemma 13: In a SLCT T , if x_i and x_j are two distinct independent variables, then any edge out of x_i does not occur in any path containing x_j .

Proof: Since x_i and x_j are both independent, x_i and x_j are of indegree 0. It follows that the source of any path containing x_i must be x_i , since every other vertex in the path has an edge into it. Similarly, the source of any path containing x_j is x_j . If an edge A from x_i were on a path containing x_j , then both x_i and x_j would have to be the source of that path. But that is impossible, since each path has a unique source, and the two variables are distinct by supposition. \square

Definition 14: Given a SLCT $\langle D, (\Omega, P), R, X, L \rangle$:

Let P_{ij} be the set of paths from x_i to x_j .

Let P_{ij}^n be the set of paths from x_i to x_j of length n .

Let $P_{ij}^{\leq n}$ be the set of paths from x_i to x_j of length less than or equal to n .

Let P_{ij}^0 be the set of open paths from x_i to x_j .

Definition 15: Given a SLCT $\langle D, (\Omega, P), R, X, L \rangle$, a **first level expansion** of a dependent variable x_j is the expression on the right side of the equation for x_j in X .

A **n level expansion** of an independent variable x_j is x_j . An **$n+1$ level expansion** of a dependent variable x_j is the expression resulting from substituting for each dependent variable x_i in the n^{th} level expansion of x_j , the first level expansion of x_i . A **full expansion** of a variable is an infinite level expansion of that variable.

Note that in the case of a theory with an acyclic graph, there always exists a finite n such that an n level expansion of a variable x is identical to any expansion level greater than n .

The following example illustrates these concepts.

$$x_3 = 1 * x_1 + 2 * x_2$$

$$x_4 = 3 * x_3 + 4 * x_1$$

$$x_5 = 5 * x_4 + 6 * x_1$$

A first level expansion of x_5 is simply

$$5 * x_4 + 6 * x_1$$

A second level expansion of x_5 is

$$5 * (3 * x_3 + 4 * x_1) + 6 * x_1$$

A third level expansion of x_5 is

$$5 * (3 * (1 * x_1 + 2 * x_2) + 4 * x_1) + 6 * x_1$$

Any higher level expansions of x_5 will be exactly the same. Hence, the last expression is also a full expansion of x_5 .

It is possible for an expression to be equal to an n level or full expansion of a variable x without being an n level or full expansion of x . For example, the following expressions are equal to a full expansion of x_5 , but are not themselves full expansions of x_5 .

$$x_5 = 41 * x_1 + 30 * x_2$$

$$x_5 = 15 * x_1 + 30 * x_2 + 20 * x_1 + 6 * x_1$$

The first expression is a modification of the complete expansion of x_5 which is a sum of terms; each combination of variables to a given power occurs in at most one term; in this case all of the x_1 terms are collected together and all of the x_2 terms are collected together.

The second expression is a modification of the complete expansion of x_5 in which the distributive law has been repeatedly applied to the complete expansion until it is a sum of terms each of which contains no additions, and in which none of the additions have actually been carried out (that is the additions in $15 * x_1 + 20 * x_1 + 6 * x_1$ have not been carried out to form $41 * x_1$). There are as many occurrences of each variable in the second expression as there are in the fully expanded form.

These two particular forms which are equal to expansions of a variable play an important

part in the proofs that follow. Their definitions are given below.

Definition 16: A **collected** n level (full) expansion of a variable x is an expression equal to the n level (full) expansion of x , that is a sum of terms in which each combination of variables to a given power occurs in at most one term.

A **distributed** n level (full) expansion of a variable x is an expression that can be derived from an n level (full) expansion of x by repeatedly distributing multiplication over addition until a sum of terms results in which each term contains no additions, and no additions of terms are carried out.

In what follows we shall speak not only of expansions of variables, but also of distributed or collected expansions of variances, covariances, and products of covariances of variables. These are all derived from the above concepts by performing the corresponding distributed or collected expansions on the variables occurring in the variances or covariances, and then applying lemma 8 or lemma 9.

Note that in a distributed or collected full expansion of a variance or a covariance, the variables occurring in each covariance will either be identical or independent; hence all of the non-zero covariances become variances.

Definition 17: Two terms are **identically equal** iff they are equal for all values of the equation coefficients.

x_i is a **direct descendant** of x_j iff x_j is into and adjacent to x_i . DD_i denotes the set of direct descendants of x_i .

Henceforth, when we write "equal" or "=" in reference to products of labels of paths we will mean identically equal unless we explicitly state otherwise. Similarly, "not equal" or "≠" in reference to products of labels of paths will mean not identically equal unless we explicitly state otherwise.

Lemma 18: Given an SLCT $\langle D, (\Omega, P), R, X, L \rangle$, the coefficient a_i of the variable x_i in a collected n level expansion of a dependent variable x_j , where $x_i \neq x_j$, is

$$\sum_{p \in P_{ij}^n} \prod_{e \in p} L(e) \quad \text{if } x_i \text{ is dependent}$$

$$\sum_{p \in P_{ij}^{\leq n}} \prod_{e \in p} L(e) \quad \text{if } x_i \text{ is independent}$$

Proof: The proof will be by induction on the expansion-level of x_j . For the remainder of the proof it is understood that all expansions are collected

expansions.

Base case: Since P_{ij}^1 is equal to $P_{ij}^{\# \leq 1}$, if $i \neq j$, it suffices to show that in a first level expansion of x_j , the coefficient a_i of x_i is equal to

$$\sum_{p \in P_{ij}^1} \prod_{e \in p} L(e)$$

There can be at most one path in P_{ij}^1 . If there is one path it consists of the path consisting of the single edge from x_i to x_j . The label of this edge is the coefficient of x_i in the equation for x_j in X . This coefficient is by definition, the coefficient of x_i in the first level expansion of x_j . If there are no paths in P_{ij}^1 then the sum is by convention equal to 0, which is the coefficient of x_i in the first level expansion of x_j if they are not connected by an edge.

Induction: Suppose that in the n^{th} level expansion of x_j , the coefficient of an independent variable x_i is equal to

$$\sum_{p \in P_{ij}^n} \prod_{e \in p} L(e) \quad \text{if } x_i \text{ is dependent}$$

$$\sum_{p \in P_{ij}^{\leq n}} \prod_{e \in p} L(e) \quad \text{if } x_i \text{ is independent}$$

It will be shown that in the $n+1^{\text{th}}$ level expansion of x_j , the coefficient of a variable x_i is equal to

$$\sum_{p \in P_{ij}^{n+1}} \prod_{e \in p} L(e) \quad \text{if } x_i \text{ is dependent}$$

$$\sum_{p \in P_{ij}^{\leq n+1}} \prod_{e \in p} L(e) \quad \text{if } x_i \text{ is independent}$$

There are two cases. Suppose first that x_i is a dependent variable. In the n^{th} level expansion of x_j , the coefficient of any variable which is not in DD_i will not be a factor in the coefficient of x_i in the $n+1^{\text{th}}$ level expansion of x_j . This is because when the n^{th} level expansion of x_j is first-level expanded to produce the $n+1^{\text{th}}$ level expansion of x_j , any term not in DD_i will be expanded into a term which does not have x_i as a factor.

On the other hand, the coefficients of any direct descendants of x_i that appear in the n^{th} level expansion of x_j will be factors in the coefficient of x_i at the $n+1^{\text{th}}$ level expansion of x_j . This is because the first-level expansion of any

direct descendant of x_i will replace that direct descendant with a sum of variables, one of which is x_i .

Take an arbitrary random variable x_z which is a direct descendant of x_i and appears in the n^{th} level expansion of x_j . By the induction hypothesis, the coefficient of x_z is

$$\sum_{p \in P_{zj}^n} \prod_{e \in p} L(e)$$

It follows that after a 1^{st} level expansion of an n^{th} level expansion of x_j , the coefficient of x_i is equal to

$$\sum_{x_z \in DD_i} \left(\left(\sum_{p \in P_{zj}^n} \prod_{e \in p} L(e) \right) * \left(\sum_{p' \in P_{iz}^1} \prod_{e \in p'} L(e) \right) \right) \tag{1}$$

(Each term in this formula is the result of taking the product of coefficients along a path of length n from a direct descendant x_z of x_i to x_j , and multiplying it by the coefficient of an edge from x_i to x_z .)

It must be shown now that formula 1 is equal to

$$\sum_{p'' \in P_{ij}^{n+1}} \prod_{e \in p''} L(e) \tag{2}$$

We will show that formula 1 is equal to 2 by showing that there is a 1-1 function mapping terms in 1 onto equal terms in 1.

Each term in formula 1 is the result of taking the product of labels along a path of length n through a direct descendant x_z of x_i to x_j and multiplying it by the coefficient of an edge from x_i to x_z . The result is a product of labels along a path of length $n+1$ from x_i to x_j , which corresponds uniquely and is equal to a term in formula 2. (See 10-8.)

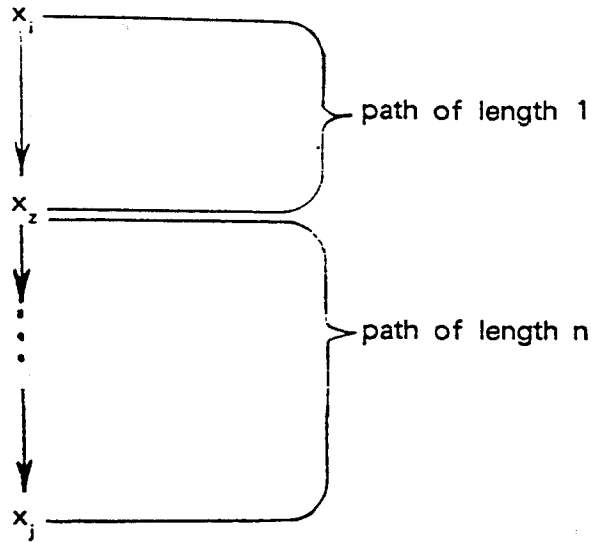


Figure 10-8:

Similarly, every term in formula 2 uniquely corresponds to and is equal to a term in formula 1. This follows from the fact that every path in P_{ij}^{n+1} can be broken into two subpaths. One subpath is a path from x_i to x_z of length 1, where $z \in DD_{ij}$; the other is a path of length n from x_z to x_j . The product of labels of edges of the path from x_z to x_j uniquely corresponds to and is equal to a term in

$$\sum_{p \in P_{zj}^n} \prod_{e \in p} L(e)$$

The label of the edge from x_i to x_z uniquely corresponds to and is equal to a term in

$$\sum_{p' \in P_{iz}^1} \prod_{e \in p'} L(e)$$

Hence the product of the labels of the edges of the two paths uniquely corresponds to and is equal to a term in formula 1.

The second case occurs if x_i is an independent variable. The proof is nearly the same. The contribution to the coefficient of x_i from paths of length $n+1$ is still

$$\sum_{p'' \in P_{ij}^{n+1}} \prod_{e \in p''} L(e)$$

However, there is now also a contribution to the coefficient of x_i from paths of length n or less. This is because the 1st level expansion of a n^{th} level expansion does not change any term in which an independent variable is a factor. The coefficient of x_i in the n^{th} level expansion of x_j is equal to

$$\sum_{p \in P_{ij}^{\leq n}} \prod_{e \in p} L(e)$$

by the induction hypothesis. Hence, the total coefficient of x_i in the $n+1^{\text{th}}$ level expansion of x_j is equal to

$$\sum_{p' \in P_{ij}^{n+1}} \prod_{e \in p'} L(e) + \sum_{p \in P_{ij}^{\leq n}} \prod_{e \in p} L(e)$$

Since

$$P_{ij}^{\leq n+1} = P_{ij}^{n+1} \cup P_{ij}^{\leq n}$$

it is clear that if x_i is an independent variable then its coefficient in an n^{th} level expansion of x_j is

$$\sum_{p \in P_{ij}^{\leq n+1}} \prod_{e \in p} L(e)$$

□

Lemma 19: Given a SLCT $\langle D, (\Omega, P), R, X, L \rangle$, the coefficient a_i of the variable x_i in a collected full expansion of a dependent variable x_j is equal to

$$\sum_{p \in P_{ij}} \prod_{e \in p} L(e)$$

Proof: This follows immediately from lemma 18. □

Theorem 20: Any two SLCTs $T = \langle D, (\Omega, P), R, X, L \rangle$ and $T' = \langle D', (\Omega', P'), R', X', L' \rangle$ with isomorphic graphs strongly imply isomorphic sets of tetrad equations.

Proof: Suppose $\langle D, (\Omega, P), R, X, L \rangle$ strongly implies a set of tetrad equation. We will show that an arbitrary $\langle D', (\Omega', P'), R', X', L' \rangle$ where D and D' are isomorphic strongly implies an isomorphic set of tetrad equations.

Since D is isomorphic to D' , there is a function $Z: R \rightarrow R'$. For each $r \in R$ let $Q(r) = Z(r)$. We will show that Q is the function mapping $R \rightarrow R'$

establishing the isomorphism between the set of tetrad equations.

Since D is isomorphic to D' it follows that for every equation in X of the form

$$x_i = \sum a_{ij} x_j$$

there is an equation in X' of the form

$$Q(x_i) = \sum a'_{ij} Q(x_j)$$

The only differences between the equations of X and the equations of X' are that the values of the coefficients are different, and the random variables appearing in them are different. Since a tetrad equation is strongly implied by a theory iff it is implied for all values of the coefficients, the difference in the values of the coefficients between the equations of X and the equations of X' is irrelevant to which tetrad equations are strongly implied. We will now show that which random variables occur in the equations is also irrelevant to which tetrad equations are strongly implied.

Throughout the rest of this proof we assume that all expansions are distributed expansions.

The product of two covariances of T when written in fully expanded form is a sum of terms each of which is of the form

$$c_t * \text{var}(x_i) \text{var}(x_j)$$

where each c_t is equal to a product of equation coefficients of T raised to some positive power.

The product of the corresponding two covariances of T' when written in fully expanded form is a sum of terms each of which is of the form

$$c'_t * \text{var}(Q(x_i)) \text{var}(Q(x_j))$$

where a_{ij} raised to some power w is a factor of c_t iff the corresponding constant a'_{ij} raised to the same power w is a factor of c'_t .

Since the tetrad equality holds for all values of the equation coefficients, it follows that there is a 1-1 function that maps terms on the l.h.s to terms on the r.h.s. which have identically equal term coefficients; i.e., the r.h.s. must have a term coefficient that is a product of the same equation coefficients raised to the same powers. That is, for each term on the right side of the equation of the form

$$c_t * \text{var}(x_i) \text{var}(x_j)$$

there is a term on the left side of the equation of the form

$$c_u * \text{var}(x_k) \text{var}(x_z)$$

where $c_t = c_u$ for all values of the equation coefficients. Similarly, we know that $c_t' = c_u'$ for all values of the equation coefficients of T' .

There is only way that the tetrad equation could be strongly implied by T while the corresponding tetrad equation is not strongly implied by T' , since the equality of the term coefficients is an identity in each theory. That is if

$$\text{var}(x_i) * \text{var}(x_j) = \text{var}(x_k) * \text{var}(x_z)$$

but

$$\text{var}(Q(x_i)) * \text{var}(Q(x_j)) \neq \text{var}(Q(x_k)) * \text{var}(Q(x_z))$$

(Remember that here "=" means identically equals, and "≠" means not identically equal.)

But we will now show that this cannot happen. We will show that if the term coefficients of terms on each side of the equation are equal then $\{x_i, x_j\} = \{x_k, x_z\}$, which implies that the equation is an identity in each theory.

The proof is by contradiction. Suppose that c_t and c_u are products of the same equation coefficients raised to the same powers, but $\{x_i, x_j\} \neq \{x_k, x_z\}$.

Suppose without loss of generality that i is not equal to k or z . It follows from lemma 13 that x_i does not occur on any path containing x_k or x_z . But according to lemma 18 and lemma 8 one factor in the coefficient c_t of

$$c_t * \text{var}(x_i) \text{var}(x_j)$$

is the product of all labels of edges in a path with source x_i . Hence, the label of an edge out of x_i must appear as a factor in the term coefficient c_t . The coefficient c_u in the term

$$c_u * \text{var}(x_k) \text{var}(x_z)$$

is a product of labels of edges on paths with source x_k or x_z . It follows then that the label of an edge from x_i cannot appear as a factor in c_u since no edge from x_i occurs in any path from x_k or x_z . It follows then that c_t does not consist of the same coefficients raised to the same powers as c_u . This contradicts our assumption. \square

In the following theorem, " σ " signifies the standard deviation.

Theorem 21:

If $\sigma_u, \sigma_v, \sigma_w$, and $\sigma_x \neq 0$

then

$$\text{cov}(u,v) \text{cov}(w,x) = \text{cov}(u,x) \text{cov}(v,w)$$

iff

$$\rho_{uv}\rho_{wx} = \rho_{ux}\rho_{vw}$$

Proof: Suppose that $\text{cov}(u,v) \text{cov}(w,x) = \text{cov}(u,x) \text{cov}(v,w)$. Divide each side of the equation by $(\sigma_u \sigma_v \sigma_w \sigma_x)$. But

$$\rho_{uv}\rho_{wx} = \text{cov}(u,v) \text{cov}(w,x) / (\sigma_u \sigma_v \sigma_w \sigma_x)$$

$$\rho_{ux}\rho_{vw} = \text{cov}(u,x) \text{cov}(w,v) / (\sigma_u \sigma_x \sigma_w \sigma_v)$$

Hence,

$$\rho_{uv}\rho_{wx} = \rho_{ux}\rho_{vw}$$

The only if clause can be proved similarly. \square

The corresponding theorems can similarly be demonstrated for the other independent tetrad equations among u , v , w , and x .

Definition 22: Given a SLCT $T = \langle D, (\Omega, P), R, X, L \rangle$, let T_{ij} equal the set of all treks from x_i to x_j .

An edge e is in trek t iff it is in at least one of the paths in t .

The length of t is the sum of the lengths of the paths in t .

The following theorem states the trek rules for calculating covariances.

Theorem 23: Given a SLCT $T = \langle D, (\Omega, P), R, X, L \rangle$, the covariance of any two distinct variables x_i and x_j ($\text{cov}(x_i, x_j)$) is equal to

$$\sum_{t \in T_{ij}} \left(\left(\prod_{e \in t} L(e) \right) * \text{var}(S(t)) \right) \quad (3)$$

where $S(t)$ is the source of the trek.

Proof: We assume that all expansions in this proof are distributed expansions.

Let U be the set of indices of independent variables.

It follows from lemma 8 and lemma 19 that

$$\text{cov}(x_i, x_j) =$$

(4)

$$\sum_{u \in U} \left(\left(\sum_{p \in P_{uj}} \prod_{e \in p} L(e) \right) \left(\sum_{p \in P_{ui}} \prod_{e \in p} L(e) \right) * \text{var}(x_u) \right)$$

We will show that when each variance term in formula 3 is fully expanded, then formula 3 is equal to formula 4. From now on, formula 3 will refer to the fully expanded form of the formula.

The general strategy is to show how to associate with each term in formula 3 a unique term in formula 4, and vice-versa, and then to show that the term in formula 3 is identically equal to its associated term in formula 4.

First we will show how to associate a term in formula 3 with a unique term in formula 4; then we will show that the two associated terms are equal.

Take an arbitrary term in formula 3. Let us call $S(t) x_i$. It follows from lemma 9 and lemma 19 that

$$\text{var}(x_i) = \sum_{u \in U} \left(\left(\sum_{p \in P_{ut}} \prod_{e \in p} L(e) \right)^2 * \text{var}(x_u) \right)$$

Also, if t is a trek, then

$$\prod_{e \in t} L(e)$$

can be written as

$$\prod_{e \in t_1} L(e) \prod_{e \in t_2} L(e)$$

It follows that one term in formula 3 for a given trek t and independent variable x_u is of the form

(5)

$$\prod_{e \in t_1} L(e) \prod_{e \in p} L(e) \prod_{e \in t_2} L(e) \prod_{e \in p'} L(e) * \text{var}(x_u)$$

where p and p' are paths from x_u to x_i . Assume without loss of generality that t_1 is the path from x_i to x_j and t_2 is the path from x_i to x_j . There are two cases to consider.

The first case is if p is equal to p' . If p is equal to p' , then there is one term of the form 5. Associate this term with the term in 4 which consists of the product of labels of edges in the paths consisting of the concatenation of t_1 with p and t_2 with p' . The result of the pair of concatenations is a path from

x_u to x_i and a path from x_u to x_j . Hence the associated term clearly occurs in 4 and is equal to the term in 3. (See 10-9.)

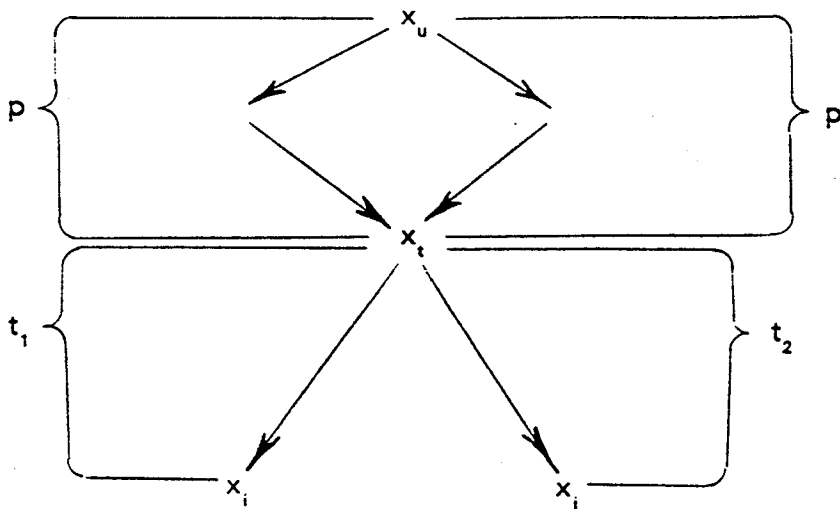


Figure 10-9:

In the second case, where $p \neq p'$, the proof is the same except that there are two terms of the form of 5. Associate one of the terms with the term in 4 that consists of the product of labels of edges of the paths that are the concatenation of t_1 with p and t_2 with p' ; the other term in 3 is associated with term in 4 that consists of the product of labels of edges of the paths that are the concatenation of t_1 with p' and t_2 with p . Again both of these terms clearly occur in 4 and are equal to their associated terms in 3.

Now we will show that an arbitrary term in formula 4 can be uniquely associated with a term in formula 3 which is equal to it. A single term in formula 4 for a given u is the product of the labels of edges in a path p from x_u to x_i times the product of the labels of the edges in a path q from x_u to x_j . Let x_t be the last point of intersection of these two paths. It is clear that x_t is the source of a trek from x_i to x_j , since the paths from x_t to x_i and from x_t to x_j don't intersect except at x_t . We can break p into two parts; p_1 which is the subpath from x_u to x_t , and t_1 which is the path from x_t to x_i . Similarly, we can break q into two parts, q_1 from x_u to x_t and t_2 from x_t to x_j . Hence the arbitrary term in formula 4 can be written as

$$\left(\prod_{e \in t_1} L(e) \right) \left(\prod_{e \in p_1} L(e) \right) \left(\prod_{e \in t_2} L(e) \right) \left(\prod_{e \in q_1} L(e) \right)$$

(6)

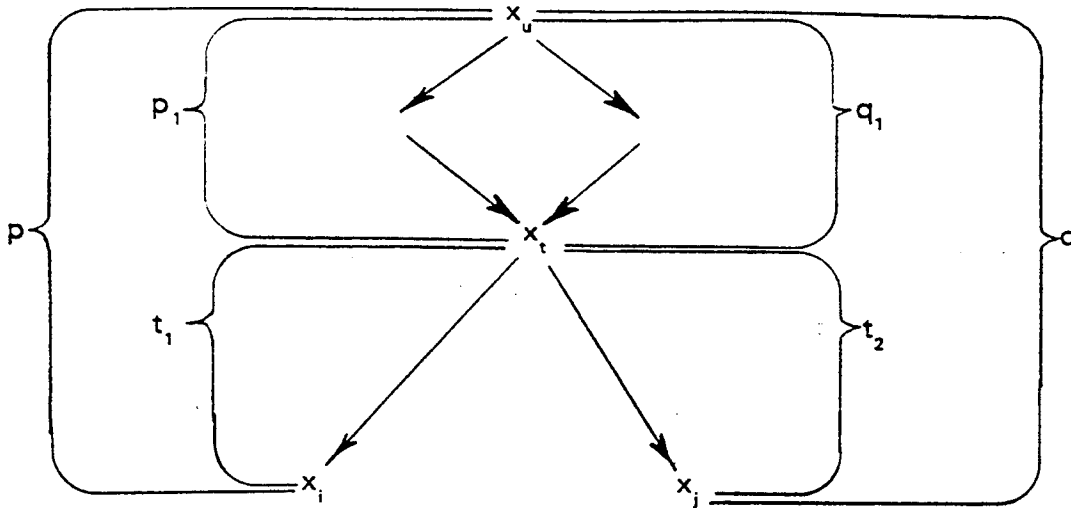


Figure 10-10:

Since t is the source of a trek from x_i to x_j , there is a term in the unexpanded form of formula 3 of the form

$$\left(\prod_{e \in t} L(e) \right) * \text{var}(x_i) \tag{7}$$

When $\text{var}(x_i)$ is fully-expanded

$$\text{var}(x_i) = \sum_{u \in U} \left(\left(\sum_{p \in P_{ut}} \prod_{e \in p} L(e) \right)^2 * \text{var}(x_u) \right)$$

When this is substituted in to formula 7, the result is

$$\prod_{e \in t} L(e) * \sum_{u \in U} \left(\left(\sum_{p \in P_{ut}} \prod_{e \in p} L(e) \right)^2 * \text{var}(x_u) \right) \tag{8}$$

which is a sum of terms.

Since both p_1 and q_1 are in P_{ut} it follows that one of the terms in formula 8 is just

$$\left(\prod_{e \in t_1} L(e) \right) \left(\prod_{e \in p_1} L(e) \right) \left(\prod_{e \in t_2} L(e) \right) \left(\prod_{e \in q_1} L(e) \right) \tag{9}$$

But this is just the arbitrary term of formula 4.

(Note that if p_1 does not equal q_1 , then there will be two terms of the form 6. However, there will also be two terms of the form 9 occurring in 3. Hence

distinct terms in 4 can still be mapped onto distinct terms in 3.)

It follows that there is a 1-1 function mapping each term in 3 onto an equal term in 4, so the two formulas are equal. \square

Tetrad does not correctly calculate the tetrad equations strongly implied by every graph, as the following example will show.

It is possible that a tetrad equation holds among four variables, even though in the trek form of the equations for the covariances, there is a term on the left side of the equation which has a coefficient not equal to the coefficient of any term on the right side of the equation. Consider example 1 again.

We saw that

$$\text{cov}(x_5, x_6) * \text{cov}(x_3, x_4) =$$

$$\text{cov}(x_3, x_6) * \text{cov}(x_4, x_5) =$$

$$a_{31} a_{62} a_{41} a_{52} a_{21}^2 * \text{var}(x_1)$$

However, when the covariances are written in trek form the left side of the equation takes the form

$$\text{cov}(x_5, x_6) * \text{cov}(x_3, x_4) =$$

$$a_{62} a_{41} a_{52} a_{31} \text{var}(x_2) * \text{var}(x_1)$$

Written in this form the coefficient on the right side of the equation ($a_{21}^2 a_{62} a_{41} a_{52}$) does not equal the coefficient on the left side of the equation ($a_{31} a_{62} a_{41} a_{52}$).

When TETRAD determines whether or not a tetrad equation among four variables holds, it compares only the coefficients and ignores the variances. If there is a term on either side of the equation which does not have a corresponding term on the other side of the equation with an equal coefficient then it concludes that the tetrad equation does not hold. For that reason, TETRAD would have concluded that the tetrad equation described above was not strongly implied. However, it can be shown that on a restricted set of graphs (including all of the ones of interest to social scientists) comparing the coefficients of the trek forms of products of covariances, while ignoring the variances they are multiplied by, will correctly compute which tetrad equations are strongly implied.

In effect, TETRAD assumes that in every graph, every non-independent variable has associated with it a unique error term that is uncorrelated with any other error term in the graph. Tetrad does not calculate the tetrad equations strongly implied by the graph

that the user enters; it attaches to every non-independent variable a unique error term uncorrelated with every other error term.

If the user entered the graph of example 1, the actual graph that Tetrads would perform its calculations on, is depicted below.

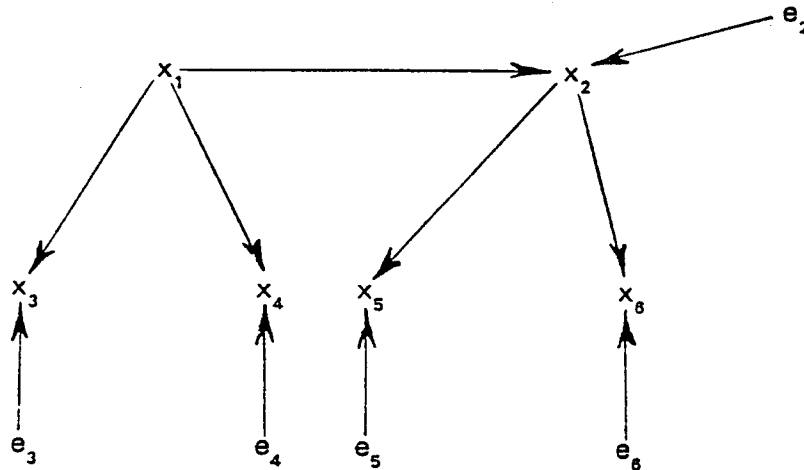


Figure 10-11:

A graph is Tetrads-compatible iff every vertex not of indegree 0 is adjacent to a vertex of indegree 0.

The following theorem shows that the method that Tetrads uses to compute which tetrad equations are strongly implied is correct if the digraph D of the theory is tetrad-compatible.

Given a SLCT $T = \langle D, (\Omega, P), R, X, L \rangle$, when the covariance of two variables is written in fully-expanded form and there are cycles in D , the expression can be a sum of an infinite number of terms. Mason has shown how to reduce this sum of an infinite number of terms into an equal sum of a finite number of terms.³⁵ Mason's rule will be used in the proof of the correctness of the Tetrads algorithm. The following definitions are used in the statement of Mason's rule.

Definition 24: An open path is a path which contains no cycles. A loop is a path in which the head of the source equals the tail of the sink, and no edge occurs in the path twice.

A loop touches an open path iff the loop and the open path intersect in a vertex which is not the source of the open path. A loop touches a loop iff

³⁵ [40], pp. 55-88

the two loops intersect.

A loop is relevant to the variables x_i and x_j iff it either touches some open path between x_i and x_j or it touches another loop relevant to the variables x_i and x_j .

RL_{ij} is the set of loops relevant to the variables x_i and x_j .

Theorem 25: The coefficient of x_i in the fully expanded equation for x_j is equal to

$$\frac{\left(\sum_{p \in P_{ij}^0} \prod_{e \in p} L(e) \right) \prod_{z \in RL_{ij}} \left(1 - \prod_{e \in z} L(e) \right) *}{\prod_{z \in RL_{ij}} \left(1 - \prod_{e \in z} L(e) \right)}$$

where * is a special operation in which multiplication is carried out before division, monomials are deleted if their factors are products of labels of edges of touching paths, and the division is carried out only after the deletions have been performed.

Definition 26: Let Ind be the set of independent variables in a theory. If $cov(x_i, x_j)$ is written in the form

$$\sum_{x_k} \left(\text{var}(x_k) \frac{\left(\sum_{p \in P_{ki}^0} \prod_{e \in p} L(e) \right) \prod_{z \in RL_{ki}} \left(1 - \prod_{e \in z} L(e) \right) *}{\prod_{z \in RL_{ki}} \left(1 - \prod_{e \in z} L(e) \right)} \frac{\left(\sum_{p \in P_{kj}^0} \prod_{e \in p} L(e) \right) \prod_{z \in RL_{kj}} \left(1 - \prod_{e \in z} L(e) \right) *}{\prod_{l \in RL_{kj}} \left(1 - \prod_{e \in l} L(e) \right)} \right)$$

it is said to be in **loop form**. If each x_k is an independent variable then the covariance is in **completely expanded loop form**. If each x_k is the source of a trek between x_i and x_j then it is in **loop trek form**.

An **open trek** is a trek in which each path in the trek has no cycles.

Let T_{ij}^0 be the set of all open treks between x_i and x_j .

An expression for a covariance is in **TETRAD trek form** iff it is of the form

$$\sum_{t \in T_{ij}^0} \left(\left(\prod_{e \in t} L(e) \right) * \text{var}(S(t)) \right)$$

where $S(t)$ is the source of trek t , and none of the additions between terms are carried out. (This is the form in which the TETRAD program calculates covariances, except that it leaves off the variance factors).

The coefficient of a term in TETRAD trek form is a **TETRAD trek coefficient**.

It is clear that a theory T strongly implies a given tetrad equation iff when the tetrad equation is written in completely expanded loop form, for every monomial on the l.h.s there is a unique equal monomial on the r.h.s.

Given a SLCT $T = \langle D, (\Omega, P), R, X, L \rangle$ with no loops in D , every monomial in the loop form of the covariance is equal to a monomial in the distributed fully expanded form of the covariance. If there are loops in D , then there may be monomials in the distributed loop form of the covariance which are not equal to any monomial in the distributed fully expanded form of the covariance (although of course the sum of the monomials in the loop form equals the sum of the monomials in fully expanded form).

Lemma 27: Given an SLCT with a tetrad-compatible graph, if when the variables in a tetrad equation are written in TETRAD trek form there is a 1-1 function f mapping each monomial P on the l.h.s of the equation to an identically equal monomial P' on the r.h.s, then the tetrad equation is strongly implied by the SLCT.

Proof: There are two cases to consider. Either the graph of the theory contains loops, or it doesn't. First we will consider the case in which the graph does not contain loops.

Each monomial on each side of a tetrad equation in TETRAD trek form consists of a product of labels of edges occurring in two open treks, times the variances of the sources of those open treks. Suppose that when the tetrad equation is written in TETRAD trek form, f maps a term P on the l.h.s to a term P' on the r.h.s. with an equal term coefficient. Let x_i and x_j be the sources of the treks of the monomial on the l.h.s of the equation, and x_k and x_2 be the sources of the treks of the monomial on the r.h.s of the equation. We will now show that $\{x_i, x_j\}$ is equal to $\{x_k, x_2\}$. The proof will be a reductio.

Suppose, contrary to our hypothesis, and without loss of generality that x_i is not equal to x_k or x_2 . Since x_i is the source of a trek, it follows that the labels of two edges from x_i occur in the TETRAD term coefficient of the monomial on the l.h.s. From the fact that the TETRAD term coefficient of the

monomial on the l.h.s equals the TETRAD term coefficient of the monomial on the r.h.s it follows that both edges from x_i must occur in paths with sources that are either x_k or x_z . The edges from x_i must lie in different paths; if they were in the same path, then x_i would occur twice in that path. But since the paths are acyclic that can't happen. Also, the edges from x_i must lie in different treks; if they were in the same trek, then that the paths making up that trek would have to intersect in x_i . But if the paths intersect in x_i then x_i would have to be the source of the trek, contrary to our assumption. It follows that there must be a path from x_k to x_i and a path from x_z to x_i .

From the fact that the TETRAD term coefficient of the monomial on the l.h.s equals the TETRAD term coefficient of the monomial on the r.h.s it follows that both edges from x_k and x_z must occur in paths with sources that are either x_i or x_j . Since the graph is acyclic, and x_i lies on a path from x_k and on a path from x_z , it cannot be the case that x_k or x_z lies on a path beginning at x_i . Therefore both edges from x_k and both edges from x_z lie on paths with source x_j . We will now show that this leads to a contradiction.

It is not possible for both edges from x_k to lie on the same path from x_j since the paths are acyclic. If the edges from x_k lie on different paths from x_j then the two paths from x_j intersect at x_k . If x_j is the source of the trek however, this is possible only if x_j is equal to x_k . By the same argument, x_j is equal to x_z . If x_j equals x_k then x_j does not equal x_i , since x_i does not equal x_k . But since the paths from x_j are acyclic, then the labels of edges from x_j appear in the TETRAD term coefficient of the monomial on the l.h.s. exactly twice (since x_j does not equal x_i), whereas there are exactly four labels of edges from x_k and x_z (and hence from x_i) appearing in the TETRAD term coefficient of the monomial on the r.h.s. It follows that the power of the label of some edge in the TETRAD term coefficient of the monomial on the l.h.s does not equal the power of the label of some edge in the TETRAD term coefficient of the monomial on the r.h.s. This contradicts our assumption.

The conclusion is that if the TETRAD term coefficient of the monomial on the l.h.s equals the TETRAD term coefficient of a monomial on the r.h.s then $\{x_i, x_j\}$ is equal to $\{x_k, x_z\}$, where x_i and x_j are the sources of the treks of the monomial on the l.h.s of the equation, and x_k and x_z are the sources of the treks of the monomial on the r.h.s of the equation. But if the sources are identical, then any full expansion of the variances of the sources will also be identically equal. Since the coefficients of the monomials are equal, and the expansions of the variances of the sources are equal, it follows that when fully expanded the entire TETRAD terms are identically equal. Since this is true for any two monomials on the l.h.s and r.h.s with TETRAD term coefficients that are equal, it follows that the tetrad is strongly implied by the theory.

We now consider the case in which the graph of the theory contains loops. One side of a tetrad equation is of the form

$$\text{cov}(x_i, x_j) * \text{cov}(x_k, x_z)$$

Consider a single term in the loop form of a covariance between two variables x_i and x_j . Its value depends upon 4 elements:

1. The product of the labels of coefficients in an open path P_1 connecting

a variable x_k to x_i and the product of labels of coefficients in an open path P_2 connecting x_k to x_j .

2. RL_{ki} and RL_{kj} .
3. The loops that touch P_1 and P_2 (which affect which terms are removed).
4. The variance of x_k .

We will prove that when two TETRAD term coefficients in a tetrad equation are equal then when they are expanded into completely expanded loop form, they are still equal. We will do this by showing that when two TETRAD term coefficients are equal, then each of the four elements determining the value of the loop expansion of those monomials are also equal. This will be done in two steps. First the TETRAD trek coefficient will be expanded into trek loop form; then trek loop form will be expanded into completely expanded loop form.

The products of the labels of the coefficients of the paths in each monomial are equal by supposition.

A single TETRAD term coefficient in an expansion of one side of a tetrad equation is the product of labels of edges in two open treks, one between x_i and x_j , and one between x_k and x_z . Each trek is composed of two paths with the same source. The sinks of one trek are x_i and x_j and the sinks of the other trek are x_k and x_z .

From this, and what was proved above for open treks, it follows that if the TETRAD term coefficient of a monomial in a tetrad equation on the l.h.s is equal to the TETRAD term coefficient of a monomial on the r.h.s, then the sources of the treks of the monomial on the l.h.s are identical to the sources of the treks of the monomial on the r.h.s. Since the TETRAD term coefficients of the two monomials are equal, it follows that the same edges appear the same number of times in each TETRAD term coefficient. In a given graph, which loops touch a given path is determined solely by what edges are in the path and what the source of the path is. Each TETRAD term coefficient consists of the product of labels of edges in four open paths. Since the sources of the paths of both monomials are the same, and the edges in the paths of both monomials are the same, the loops touching the paths of both monomials are the same.

Since the sources and the sinks of the paths of each of the monomials are the same, the loops relevant to the sources and the sinks are the same. It follows then that the loop forms of the TETRAD term coefficients of each monomial are equal.

Since the sources are the same on each side, we already know that the variances of the sources on each side will be equal.

When the corresponding monomials are written in trek loop form they will therefore be equal.

We still have not written the product of covariances in completely expanded loop form, however, since the sources of the treks are not necessarily independent variables. However, since the sources of the treks of each monomial are the same, it is clear that when they are written in completely

expanded form, the two monomials will still be equal. So if there is a 1-1 function mapping each TETRAD term coefficient on the l.h.s. to an equal term coefficient on the r.h.s., it follows that when written in completely expanded loop form, the product of the covariances on each side of the equation will still be identically equal. This in turn implies that the tetrad equation is strongly implied. \square

Lemma 28: If a tetrad equation is strongly implied by a SLCT with a tetrad-compatible graph then when the covariances of the variables are written in TETRAD trek from there is a 1-1 function f mapping each monomial P on the l.h.s. of the equation to an identically equal monomial P' on the r.h.s.

Proof: We will prove that if there is no 1-1 function f mapping each monomial P on the l.h.s. of the equation to an equal monomial P' on the r.h.s. of the equation, it follows that the l.h.s. does not equal the r.h.s. when written in completely expanded loop form; this implies that the tetrad equation is not strongly implied by the theory.

Suppose that when written in TETRAD trek form, the inequality is of the form

$$\sum_{i=1}^n A_i \neq \sum_{j=1}^m B_j$$

where there exists some value of i for which $A_i \neq B_j$ for all values of j . Let NE_A be the set of all i for which $A_i \neq B_j$ for all values of j . Similarly let NE_B be the set of all j for which $B_j \neq A_i$ for all values of i . After cancellation of the equal terms on each side of the equation what is left is

$$\sum_{A_i \in NE_A} A_i \neq \sum_{B_j \in NE_B} B_j$$

Now no term on the l.h.s. is equal to any term on the r.h.s. Let us expand this inequality into distributed completely expanded loop form. Each term A_i expands into a sum of terms of the form

$$\frac{A_i C_k E_z G_m}{1 - J_n}$$

where

1. A_i is the corresponding term in the Tetrad trek form of the inequality.
2. G_m is $\text{var}(x_o) * \text{var}(x_p)$ for some pair of independent variables x_o and x_p .
3. C_k is a product of labels of edges in loops or 1.
4. E_z is a product of edges along paths from independent variables x_o and x_p to the sources of the trek corresponding to A_i , or 1 if the source is an independent variable.
5. J_n is a series of products of labels of edges in loops or 0.

Similarly each B_j expands into a sum of terms each of the form

$$\frac{B_j C_q E_r G_s}{1 - J_t}$$

We will now show that

if

$$A_i C_k E_z G_m \neq B_j C_q E_r G_s$$

then

$$\frac{A_i C_k E_z G_m}{1 - J_n} \neq \frac{B_j C_q E_r G_s}{1 - J_t}$$

Suppose that

$$A_i C_k E_z G_m \neq B_j C_q E_r G_s$$

There are three cases to consider. Either both J_n and J_t are 0, one of them is 0, or neither of them is 0.

If both J_n and J_t are both 0, then clearly

$$\frac{A_i C_k E_z G_m}{1 - J_n} \neq \frac{B_j C_q E_r G_s}{1 - J_t}$$

since

$$\frac{A_i C_k E_z G_m}{1 - J_n} = A_i C_k E_z G_m$$

and

$$\frac{B_j C_q E_r G_s}{1 - J_t} = B_j C_q E_r G_s$$

If exactly one of J_n and J_t is 0, then

$$\frac{A_i C_k E_z G_m}{1 - J_n} \neq \frac{B_j C_q E_r G_s}{1 - J_t}$$

This is because one side of the equation will have a denominator of 1. The other side of the equation will have a denominator that can't be cancelled with any term in the numerator (since there are no subtractions or additions in the numerator) and is not equal to 1.

If neither of J_n and J_t are 0, it follows that

$$A_i C_k E_z G_m (J_t) \neq B_j C_q E_r G_s$$

or

$$A_i C_k E_z G_m \neq B_j C_q E_r G_s (J_n)$$

since otherwise

$$A_i C_k E_z G_m = A_i C_k E_z G_m J_n J_t$$

From this it follows that

$$A_i C_k E_z G_m (1 - J_t) \neq B_j C_q E_r G_s (1 - J_n)$$

which in turn implies

$$\frac{A_i C_k E_z G_m}{1 - J_n} \neq \frac{B_j C_q E_r G_s}{1 - J_t}$$

Hence we can ignore the J terms.

Since no equation coefficient occurs in G_m or G_s , and both $A_i C_k E_z$ and $B_j C_q E_r$ consist solely of products of labels of edges, it is easy to see that

if

$$A_i C_k E_z \neq B_j C_q E_r$$

then

$$A_i C_k E_z G_m \neq B_j C_q E_r G_s$$

Hence we shall ignore the G terms.

It is clear from Theorem 25 that one of the C_k terms must be 1. For every variable x_t of indegree not equal to 0 there is an attached variable e_t of indegree 0 connected to x_t by an edge with label 1. Hence it follows that there exists a E_z term equal to 1. Since every possible combination of C and E terms occurs in the expansion, there is a term in the expansion of A_i which is equal to A_i (if the J and G terms are ignored).

Suppose, contrary to the theorem, that the product of the two covariances on the l.h.s does equal the product of the two covariances on the r.h.s. It follows that when written in completely expanded loop form the product of the two covariances are equal. Hence each A_i must be equal to some term on the other side of the equation. If some A_i is equal to some term $B_j C_q E_r$ on the r.h.s, then $C_q E_r$ is not equal to 1 (since by supposition A_i does not equal B_j for any j).

Substitute for each occurrence of A_i on the l.h.s. the corresponding value term $B_j C_q E_r$.

Consider now the terms on the r.h.s of the form B_j . (The argument for their existence is the same as the argument for the existence of terms of the form A_i on the l.h.s.) In order for the l.h.s to be identically equal to the r.h.s., for each term of the form B_j , there must be a corresponding term on the l.h.s. which is equal to it.

The only way for a particular B_j term on the r.h.s. to be identically equal to a term on the l.h.s. is if B_j is equal to $B_k P$, where P is some product of labels of edges not equal to 1, and $j \neq k$. (This is because no B_j term appears on the l.h.s.; only terms in which B_j is multiplied by some product of labels of edges appear on the l.h.s.)

We will now show that it is impossible for every B_j to be equal to some $B_k P$, where P is not equal to 1.

In an equation $B_i = B_j P$ let B_i be the head of the equation, and B_j be the tail. Let a chain from B_i to B_j be a sequence of equations in which the head of the first equation is B_i , the tail of the last equation is B_j , and in every equation in between the tail of one equation equals the head of the next equation. A circular chain is a chain from B_i to B_i . A circular chain cannot occur in a consistent set of equations because after a series of substitutions it would be possible to derive that $B_i = B_i P$; where P is not equal to 1. The length of a chain is the number of equations in the chain. Given B_1, \dots, B_n the longest possible non-circular chain has $n-1$ equations in it.

Suppose that there are r different B terms, and that every B_j is equal to some $B_k P$, where P is not equal to 1. Then there must be a non-circular chain that can be formed in the following way. Choose an arbitrary B_i . We know that $B_i = B_m P$ for some B_m . Similarly, we know that $B_m = B_n P$ for some n , etc. But we know that given a consistent set of equations the maximum length of the non-circular chain formed is $r-1$. But that implies that there is some B_j which is not the head of an equation in the chain. But then $B_j \neq B_k P$ for any k , contrary to our assumption. This implies that there is a term on the r.h.s. that is not equal to a term on the l.h.s. of the product of the covariances when written in completely expanded loop form. This implies that the two sides are not equal, and hence the tetrad equation is not strongly implied by the theory. \square

Theorem 29: A tetrad equation is strongly implied by a SLCT with a tetrad-compatible graph iff when the covariances of the variables are written in TETRAD trek form, there is a 1-1 function f mapping each monomial P on the l.h.s of the equation to an equal monomial P' on the r.h.s.

Proof: The theorem follows immediately from lemmas 27 and 28. \square

The last theorem proved in this section states that adding edges to a model M can only defeat equations that are implied by M ; equations that are not implied by M cannot be implied by the augmented model.

Definition 30: A SLCT $M = \langle \langle R, E \rangle, (\Omega, P), R, X, L \rangle$ is a subtheory of $M' = \langle \langle R, E' \rangle, (\Omega, P), R, X', L' \rangle$ iff E is include in E' , and L' is an extension of L .

Theorem 31: If M is a subtheory of M' , then the set of tetrad equations strongly implied by M' is a subset of the set of tetrad equations implied by M .

Proof: We will show that if a tetrad equation is not strongly implied by M , then it is not strongly implied by M' .

Note first that it follows from the definition of trek that the set of treks of M is a subset of the set of treks of M' .

Take any tetrad equation A

$$\text{cov}(x_i, x_j) * \text{cov}(x_k, x_l) = \text{cov}(x_i, x_l) * \text{cov}(x_j, x_k)$$

that is not strongly implied by M . Write A in TETRAD trek form, assuming the existence of the treks of model M . The equation when written in this form will be called A_M . Since A_M is not strongly implied by M it follows that either there exists a term on the l.h.s. which is not identically equal to any term on the r.h.s., or there exists a term on the r.h.s. which is not identically equal to any term on the l.h.s. Assume without loss of generality that the latter is the case. Let T be a term which has no match on the l.h.s. of the equation.

Now consider equation A when written in TETRAD trek form, assuming the existence of the treks of model M' . The equation when written in this form will be called $A_{M'}$. Since the set of treks of M is a subset of the treks of M' , it follows that the set of terms appearing in A_M is a subset of the set of terms appearing in $A_{M'}$. In particular, the term T appears on the r.h.s. of $A_{M'}$.

Note that every term on the l.h.s. of $A_{M'}$ is either a term that appears in A_M or contains the label of an edge in $E' - E$. T is not identically equal to any term on the l.h.s. of $A_{M'}$ by supposition. Also, since T contains only labels of edges in E , it cannot be identically equal to any term that contains the label of an edge in $E' - E$. Hence, T is not equal to any term on the l.h.s. of $A_{M'}$.

10.2.3. Proof of Correctness of the Algorithm for Computing Vanishing Partial Correlations

The TETRAD program determines all of the vanishing partial correlations implied by a model. The algorithm used in the computation is just the obvious application of the theorems about vanishing partial correlations proved in Chapter 4. That is, the program determines all treks, and then for each triple of distinct variables, uses the theorems of Chapter 4 to determine whether or not a vanishing partial correlation is implied for the variables in the triple. The following sequence of lemmas leads to a proof of the graph theoretic characterization of the vanishing partial correlations implied by a graph.

In what follows, when a trek t is said to be equal to the product of two other treks p and q , this means that the product of labels of edges in t is identically equal to the result of multiplying the product of labels of edges in p by the product of labels of edges in q . $\text{Cov}(x, y)$ will be denoted by σ_{xy} , $\text{var}(x)$ will be denoted by σ_x^2 , and the standard deviation of x will be denoted by σ_x .

Lemma 32: Given a SLCT $\langle\langle R,E \rangle\rangle, (\Omega,P), R, X, L\rangle$, for all distinct vertices x,y,z in R , if the variances of all the variables are equal to 1, and $\rho_{xz,y} = 0$ for all values of the edge labels where $\rho_{xz,y}$ is defined, then,

- For all $t \in T_{xz}, y \in t$
- Either every $t \in T_{yz}$ has an edge out of y , or every $t \in T_{xy}$ has an edge out of y .

Proof: First we will show that if $\rho_{xz,y} = 0$ then for all $t \in T_{xz}, y \in t$

Assume $\rho_{xz,y} = 0$. It follows from the definition of partial correlation that if $\rho_{xz,y} = 0$ then $\rho_{xy} * \rho_{zy} = \rho_{xz}$. Hence, $\rho_{xy} * \rho_{zy} = \rho_{xz}$.

By definition,

$$\rho_{xz} = \frac{\sigma_{xz}}{\sigma_x * \sigma_y}$$

It has already been shown that σ_{xz} equals

$$\sum_{t \in T_{xz}} \left(\left(\prod_{e \in t} L(e) \right) * \sigma_{S(t)}^2 \right)$$

Similarly, σ_{xy} equals

$$\sum_{t \in T_{xy}} \left(\left(\prod_{e \in t} L(e) \right) * \sigma_{S(t)}^2 \right)$$

and σ_{yz} equals

$$\sum_{t \in T_{yz}} \left(\left(\prod_{e \in t} L(e) \right) * \sigma_{S(t)}^2 \right)$$

Since $\rho_{xy} * \rho_{zy} = \rho_{xz}$ it follows that

$$\begin{aligned} \sum_{t \in T_{xz}} \left(\left(\prod_{e \in t} L(e) \right) * \sigma_{S(t)}^2 \right) &= \\ \sum_{t \in T_{xy}} \left(\left(\prod_{e \in t} L(e) \right) * \sigma_{S(t)}^2 \right) * \sum_{t \in T_{yz}} \left(\left(\prod_{e \in t} L(e) \right) * \sigma_{S(t)}^2 \right) \end{aligned}$$

Since the variances of all of the variables are assumed to be 1, it follows that for all x,y that $\sigma_{xy} = \rho_{xy}$, and

$$\sum_{t \in T_{xz}} \left(\prod_{e \in t} L(e) \right) = \sum_{t \in T_{xy}} \left(\prod_{e \in t} L(e) \right) * \sum_{t \in T_{yz}} \left(\prod_{e \in t} L(e) \right)$$

Since every trek $t \in T_{xz}$ is a product a trek $\in T_{xy}$ and a trek $\in T_{yz}$, it follows that y occurs in every trek $t \in T_{xz}$.

Now we will show by reductio ad absurdum that if $\rho_{xz,y} = 0$ then either every $p \in T_{yz}$ has an edge out of y , or every $q \in T_{xy}$ has an edge out of y . Assume on the contrary that there exist treks $p \in T_{xy}$ and $q \in T_{yz}$ such that both p and q have no edge out of y . It follows then that any trek t that is equal to the product of p and q has no edge out of y and one of the termini of t must be y . But it was just shown that the product of any trek $\in T_{xy}$ with any trek $\in T_{yz}$ is equal to a trek $\in T_{xz}$. But it is impossible for the same trek t to terminate in x and z , and also terminate in y , since y is distinct from x and z . \square

Lemma 33: Given a SLCT $\langle\langle R,E \rangle, (\Omega,P), R, X, L \rangle$, for all distinct vertices x,y in R , if all treks $\in T_{xy}$ contain an edge out of x , then all treks $\in T_{xy}$ are paths from x to y .

Proof: Suppose that all of the treks between x and y contain an edge out of x . For each trek, there are three cases to consider: either a trek between x and y is a path from x to y , a path from y to x , or there is a pair of paths from another variable z to x and y .

First we will show by reductio ad absurdum that no trek between x and y can be a path from y to x . Assume then that every trek between x and y contains an edge out of x , but that one of the treks between x and y is a path from y to x . Such a path must contain an edge into x . We have assumed that there is an edge out of x . Since there are edges both into and out of x , and the path must terminate at x , there must be a cycle in the path. It is obvious that every cyclic path between y and x must contain an acyclic subpath from y to x . Hence, there must be an acyclic path from y to x . But this acyclic path from y to x must contain an edge into x , and hence cannot contain an edge out of x . This is contrary to our hypothesis. It follows that no trek between x and y can be a path from y to x .

Similarly, it can be shown that if there is a trek between x and y which consists of a pair of paths from a vertex z to x and y respectively, it follows that there is an acyclic path from z to x which contains no edge out of x . This would be contrary to our hypothesis. Therefore no trek between x and y consists of a pair of paths from a vertex z to x and y respectively.

Since no trek can be a path from y to x or a pair of paths from z to x and y , each trek must be a path from x to y . \square

Lemma 34: Given a SLCT $\langle\langle R,E \rangle, (\Omega,P), R, X, L \rangle$, for all distinct vertices x,y,z in R , if for all $t \in T_{xz}$ there exists a path p from y to z that is contained in t , then

- t contains a subtrek $q \in T_{xy}$
- t is equal to the product of p and q
- t and q have the same source.

Proof: If T_{xz} is empty the theorem is vacuous. Otherwise, there are three cases to consider, corresponding to the three different kinds of treks between x and z .

First, suppose that the trek t between x and z is a path from x to z , where y lies along the path. Then there is a subpath from x to y that is a trek from x to y , and a subpath from y to z that is a trek from y to z . The source of the trek between x and y is the same as the source of the trek between x and z (namely x). It is clear that t is equal to the product of p and q . (See figure 10-11).

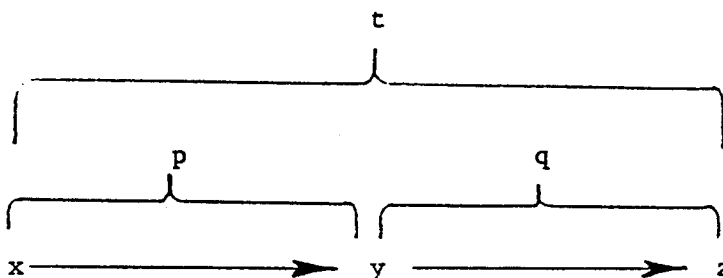


Figure 10-12: Path from x to z

In the second case, the trek t between x and z is a path from z to x . This case is not consistent with the assumptions of the theorem, however. Since every cyclic path from z to x contains an acyclic subpath from z to x , there must be an acyclic path t' from z to x . This path forms a trek between x and z , and by assumption must contain y . Hence, t' contains a path from z to y . But by assumption t' must also contain a path from y to z . This contradicts the fact that t' is acyclic. It follows that t cannot be a path from z to x . (See figure 10-12).

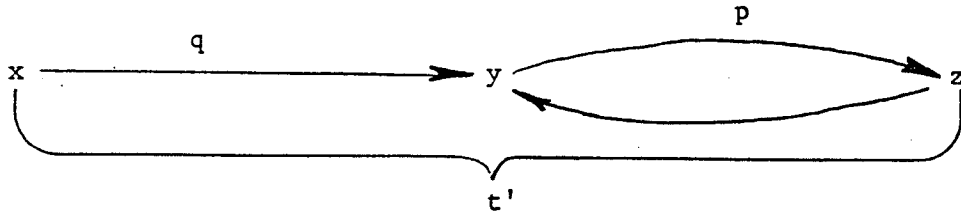


Figure 10-13: Path from z to x

In the third case, the trek t between x and z consists of a path p from $S(t)$ to x , and a path q from $S(t)$ to z . y lies on either p or q . If y lies on p , then there is a path from $S(t)$ to y , and a path from $S(t)$ to z , which form a trek between y and z . But this trek between y and z is not a path between y and z unless y is $S(t)$, in which case y also lies on q . It follows then that y must lie on q . In that case there is a trek between x and y that consists of p and the path from $S(t)$ to y . There is also a trek between y and z that is a path from y to z . The source of the trek between x and y is the same as the source of the trek between x and z (namely $S(t)$). The pair of treks between y and z , and t is obviously the product of x and y . (See figure 10-13).

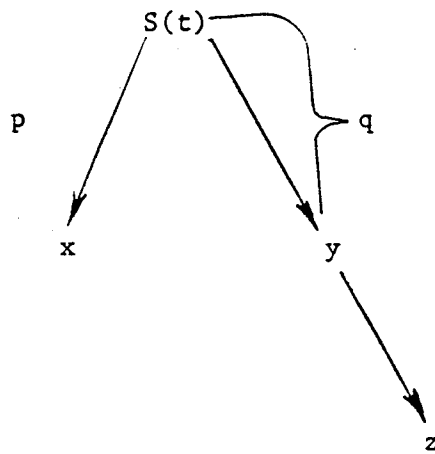


Figure 10-14: Paths from $S(t)$ to z and x

□

Lemma 35: Given a SLCT $\langle\langle R, E \rangle, (\Omega, P), R, X, L\rangle$, for all distinct vertices x, y, z in

R, if p is a path from y to z , then for each trek $q \in T_{xy}$ there exists a trek $t \in T_{xz}$ such that

- t is equal to the product of p and q
- t and q have the same source.

Proof: In every case, the trek between y and z consists of a path from y to z . There are three cases to consider for the trek between x and y .

First, suppose that q is a path from x to y . Then the concatenation of p and q is a path from x to z , which is a trek between x and z . The product of p and q is equal to a trek $t \in T_{xz}$. Furthermore, the source of the trek between x and z is the same as the source of the trek between x and y (namely x). See figure 10-14.

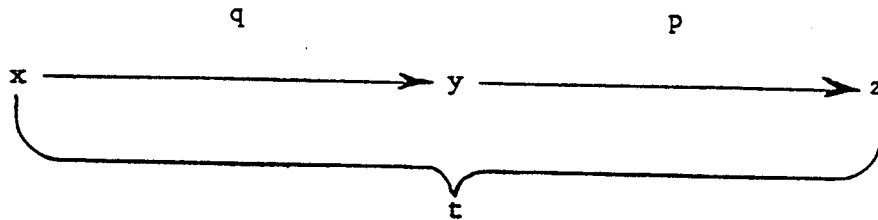


Figure 10-15: Q is a path from x to y

Second, suppose that the trek q from x to y is a path from y to x . In that case, there is a path q from y to x and a path p from y to z , which form a trek between x and z . The product of p and q is equal to a trek $t \in T_{xz}$. The source of the trek between x and z is the same as the source of the trek between x and y (namely y). See figure 10-15.

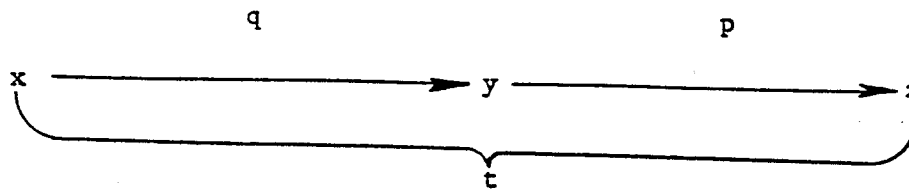


Figure 10-16: Q is a path from y to x

Third, suppose that the trek q from x to y is a pair of paths from some vertex r , to x and y respectively. In that case, the concatenation of the path from r to y , and the path from y to z is a path from r to z . Hence, there exists a path from r to x , and a path from r to z , which forms a trek between x and z . The product of p and q is equal to a trek $t \in T_{xz}$. The source of the trek between x and z is the same as the source of the trek between x and y (namely r). See figure 10-16.

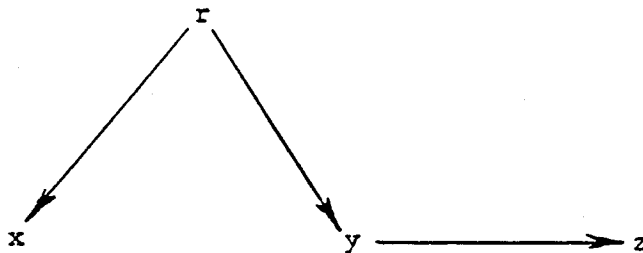


Figure 10-17: Q is a pair of paths from r to x and y

Lemma 36: Given a SLCT $\langle\langle R,E \rangle, (Q,P), R, X, L \rangle$, for all distinct vertices x,y,z in R , if

- For all $t \in T_{xz}$, $y \in t$
- Either every $t \in T_{yz}$ has an edge out of y , or every $t \in T_{xy}$ has an edge

out of y

- the variances of all variables are equal to 1

then $\rho_{xz,y} = 0$ for all values of the labels where it is defined.

Proof: Since the variances of all the variables are 1, for all x and y , $\sigma_{xy} = \rho_{xy}$.

ρ_{xy} equals

$$\sum_{t \in T_{xy}} \left(\prod_{e \in t} L(e) \right) \quad (10)$$

and $\rho_{xy} * \rho_{yz}$ equals

$$\sum_{t \in T_{xy}} \left(\prod_{e \in t} L(e) \right) * \sum_{t \in T_{yz}} \left(\prod_{e \in t} L(e) \right) \quad (11)$$

It will be assumed that both formulas 10 and 11 are written in distributed fully-expanded form.

We will first show that given that y occurs in every trek between x and z that to each term in 10 there corresponds a unique identically equal term in 11.

From lemma 33 it follows that either every trek between x and y is a path from y to x , or every trek between y and z is a path from y to z . This, together with lemma 34 implies every trek $t \in T_{xz}$ is equal to a product of treks $p \in T_{xy}$ and $q \in T_{yz}$. The product of labels of edges in t is a term in formula 10. The product of labels in the p and q form a term in formula 11. It follows that for every term in 10 there is a corresponding identically equal term in 11. Furthermore, since any two distinct treks between x and z differ in the number of occurrences of at least one edge, two distinct terms in formula 10 cannot be identically equal to the same term in formula 11.

Next we will show that given that y occurs on every trek between x and z , and that either every $t \in T_{yz}$ has an edge out of y , or every $t \in T_{xy}$ has an edge out of y , it follows that for each term in formula 11 there is a unique identically equal term in formula 10.

Assume without loss of generality that every trek between y and z has an edge out of y . By lemma 33 each trek between y and z is a path q from y to z .

Consider an arbitrary term in formula 11. It is the product of labels of edges in a trek p between x and y (which has the same edges as a path from y to x), and a trek q between y and z . From lemma 35 it follows that the product of these is equal to a trek $t \in T_{xz}$. The product of p and q form a term in 11. The product of labels of edges in t is a term in 10. Since the edges in p and q on the one hand, and t on the other hand, are identical, it follows that to the arbitrary term in 11 there is an identically equal term in 10. Furthermore, since no two terms in 11 are identically equal, it follows that no two terms in 11 are identically equal to a single term in 10. Hence, for each term in 11 there is a unique identically equal term in 10.

It follows that 11 is identically equal to 10, and hence that $\rho_{xz} = \rho_{xy} * \rho_{yz}$ for all values of the labels of the edges, as long as the variances of all of the variables are equal to 1. This implies that $\rho_{xz,y} = 0$ if $\rho_{xz,y}$ is defined. \square

Lemma 37: Given a SLCT $\langle\langle R,E \rangle, (\Omega,P), R, X, L \rangle$, for all distinct vertices x,y,z in R , $\rho_{xz,y} = 0$ (where defined) for all values of the edge labels given that the variances of all the variance are equal to 1 iff $\rho_{xz,y} = 0$ (where defined) for all values of the edge labels and all values of the variances.

Proof: The if clause is trivial.

The only if clause can be proved as follows. Assume that for all distinct vertices x,y,z that $\rho_{xz,y} = 0$ when all of the variances are 1. It follows from the definition of partial correlation that if the partial correlation is defined then $\rho_{xz,y} = 0$ iff $\rho_{xy} * \rho_{zy} = \rho_{xz}$. Hence, $\rho_{xy} * \rho_{zy} = \rho_{xz}$.

By definition,

$$\rho_{xz} = \frac{\sigma_{xz}}{\sigma_x * \sigma_y}$$

It has already been shown that σ_{xz} equals

$$\sum_{t \in T_{xz}} \left(\left(\prod_{e \in t} L(e) \right) * \sigma_{S(t)}^2 \right)$$

Similarly, σ_{xy} equals

$$\sum_{t \in T_{xy}} \left(\left(\prod_{e \in t} L(e) \right) * \sigma_{S(t)}^2 \right)$$

and σ_{yz} equals

$$\sum_{t \in T_{yz}} \left(\left(\prod_{e \in t} L(e) \right) * \sigma_{S(t)}^2 \right)$$

Since $\rho_{xy} * \rho_{zy} = \rho_{xz}$ it follows that

$$\begin{aligned} \sum_{t \in T_{xz}} \left(\left(\prod_{e \in t} L(e) \right) * \sigma_{S(t)}^2 \right) &= \\ \sum_{t \in T_{xy}} \left(\left(\prod_{e \in t} L(e) \right) * \sigma_{S(t)}^2 \right) * \sum_{t \in T_{yz}} \left(\left(\prod_{e \in t} L(e) \right) * \sigma_{S(t)}^2 \right) \end{aligned}$$

Since all of the variances of the variables are assumed to be 1, it follows that

for all x, y that $\sigma_{xy} = \rho_{xy}$, and

$$\sum_{t \in T_{xz}} \left(\prod_{e \in t} L(e) \right) = \sum_{t \in T_{xy}} \left(\prod_{e \in t} L(e) \right) * \sum_{t \in T_{yz}} \left(\prod_{e \in t} L(e) \right)$$

Clearly, this implies that

$$\frac{\sum_{t \in T_{xz}} \left(\prod_{e \in t} L(e) \right)}{\sigma_x * \sigma_z} = \frac{\sum_{t \in T_{xy}} \left(\prod_{e \in t} L(e) \right)}{\sigma_z} * \frac{\sum_{t \in T_{yz}} \left(\prod_{e \in t} L(e) \right)}{\sigma_x}$$

By lemma 32 either every $t \in T_{xy}$ or every $t \in T_{yz}$ has an edge out of y . By lemma 33 it follows that either every trek $t \in T_{yz}$ or every trek $t \in T_{xy}$ is a path with source y . Assume without loss of generality, then, that each trek between z and y has y as its source. It follows that $\sigma_{S(t)}^2$ for each source of each trek t between z and y is σ_y^2 . Hence

$$\frac{\sum_{t \in T_{xz}} \left(\prod_{e \in t} L(e) \right)}{\sigma_x * \sigma_z} = \frac{\sum_{t \in T_{yz}} \left(\prod_{e \in t} L(e) \right) * \sigma_{S(t)}^2}{\sigma_y * \sigma_z} * \frac{\sum_{t \in T_{xy}} \left(\prod_{e \in t} L(e) \right)}{\sigma_y * \sigma_x}$$

From lemma 34 it follows that each trek between x and z is the product of two subtrees between x and y , and y and z , and that the source of the trek between x and y is the same as the source between x and z . Hence we can multiply each side of the equation by the variances of the sources of their respective treks, regardless of their values, and obtain

$$\frac{\sum_{t \in T_{xz}} \left(\prod_{e \in t} L(e) \right) * \sigma_{S(t)}^2}{\sigma_x * \sigma_z} = \frac{\sum_{t \in T_{xy}} \left(\prod_{e \in t} L(e) \right) * \sigma_{S(t)}^2}{\sigma_y * \sigma_z} * \frac{\sum_{t \in T_{yz}} \left(\prod_{e \in t} L(e) \right) * \sigma_{S(t)}^2}{\sigma_x * \sigma_y}$$

Equation 12 is just the equation for $\rho_{xy} * \rho_{zy} = \rho_{xz}$. But by definition, if $\rho_{xz,y}$ is defined, this is true iff $\rho_{xz,y} = 0$. It follows then that $\rho_{xz,y} = 0$ (where defined) for all values of the edge labels given that the values of all variances are equal to 1, iff $\rho_{xz,y} = 0$ (where defined) for all values of the edge labels and all values of the variances. \square

Lemma 38: Given a SLCT $\langle\langle R,E \rangle, (\Omega,P), R, X, L \rangle$, for all distinct vertices x,y,z in R , $\rho_{xz,y} = 0$ (where defined) for all values of the edge labels and all values of the variances of the variables iff

- For all $t \in T_{xz}$, $y \in t$.
- Either every $t \in T_{yz}$ has an edge out of y , or every $t \in T_{xy}$ has an edge out of y .

Proof: It follows immediately from lemmas 32, 36, and 37. \square

Theorem 39: Given a SLCT $\langle\langle R,E \rangle, (\Omega,P), R, X, L \rangle$, for all vertices x,y,z in R , $\rho_{xz,y} = 0$ (where defined) for all values of the edge labels and all values of the variances iff

- For all $t \in T_{xz}^0$, $y \in t$.
- Either every $t \in T_{yz}^0$ has an edge out of y , or every $t \in T_{xy}^0$ has an edge out of y .

Proof: It is trivially true that if every $t \in T_{xz}$ contains y , then every $t \in T_{xz}^0$ contains y . Similarly, it is trivially true that if every $t \in T_{xy}$ has an edge out of y , every $t \in T_{xy}^0$ has an edge out of y , and if every $t \in T_{yz}$ has an edge out of y , every $t \in T_{yz}^0$ has an edge out of y .

Every cyclic path from x to y contains an acyclic path from x to y . It follows that if every $t \in T_{xz}^0$ contains y , every $t \in T_{xz}$ contains y . Similarly, if every $t \in T_{xy}^0$ has an edge out of y , then every $t \in T_{xy}$ has an edge out of y , and if every $t \in T_{yz}^0$ has an edge out of y , then every $t \in T_{yz}$ has an edge out of y .

The theorem follows from these facts and lemma 38. \square

Part III

Using TETRAD, EQS and LISREL

11. USING TETRAD WITH EQS AND LISREL

11.1. INTRODUCTION

TETRAD is a program for helping to discover **causal structure**. It will not estimate parameters in a statistical model, nor will it perform a *statistical* test of the model as a whole. Fortunately there are programs that will. The LISREL and EQS programs are among the most widely used computer packages for estimating and testing structural equation models. EQS is always, and LISREL is usually, straightforward to use on models found with TETRAD.

The appearance of the LISREL programs made available full information maximum likelihood estimation and testing. Under the assumption of normality, the program made possible the estimation and testing of a great range of structural equation models. The LISREL programs impose certain classifications of variables, however, and these classifications and the restrictions that accompany them can sometimes cause difficulties in representing causal models. In the first part of this chapter we describe these restrictions and ways to circumvent them.

The EQS program is considerably easier to use than LISREL, but the PC version can often take 15 minutes or more for a single run. If one wants to estimate and test 20 models, say, five hours might be required. In the second section of the chapter we explain how to run EQS in batch mode so that large numbers of models can be run without the attention or time of the user.

11.2. LISREL AND ITS RESTRICTIONS

The LISREL programs are described in detail in a series of manuals. We have used the *LISREL IV User's Guide* [50] and the *LISREL VI User's Guide* [52]. LISREL requires the user to classify the variables of a model according to the following scheme:

```
ksi:  exogenous and latent
eta:  endogenous and latent
      x:  measured and dependent on ksi variables
      y:  measured and dependent on eta variables
```

Each of these variable types has an associated covariance matrix:

phi: covariance matrix of error variables for ksi variables
 psi: covariance matrix of error variables for eta variables
 theta delta: covariance matrix of error variables for x variables
 theta epsilon: covariance matrix of error variables for y variables

The program permits the user to specify covariances and variances among the phi variables, the psi variables, the theta delta variables and the theta epsilon variables. In directed graph representations of a causal model, covariances of these sorts are represented as undirected (or bidirected) edges connecting error variables.

The LISREL classification of variables carries with it some restrictions on possible causal and structural representations, including the following.

1. x variables cannot depend on eta variables.
2. x variables cannot depend on other x variables.
3. y variables cannot depend on other y variables.
4. y variables cannot depend on ksi variables.
5. No error variable from one error type can covary with an error variable from any other type.

In what follows we will use the following abbreviation scheme.

ksi = K
 eta = N
 x = x
 y = y
 psi = s
 phi = p
 theta delta = td
 theta epsilon = te

To illustrate the effect of these restrictions, we show, in Figure 11-1, a simple model that is easily expressed in LISREL's formalism. Figure 11-2 shows all of the additions to this model that can be represented in that formalism. In Figure 11-3 we show a few of the

many elaborations of the original model that cannot be represented (as is) in the LISREL formalism.

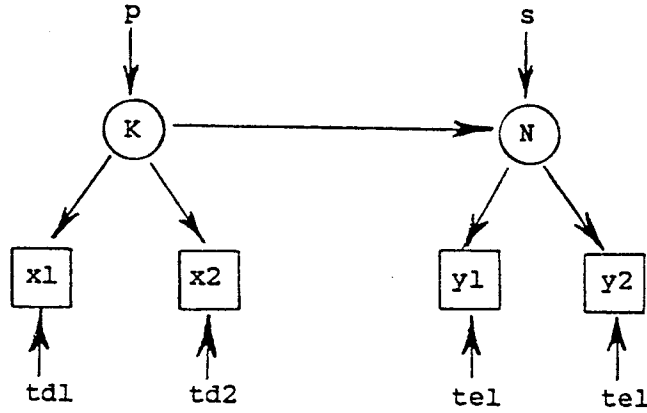


Figure 11-1: LISREL Model

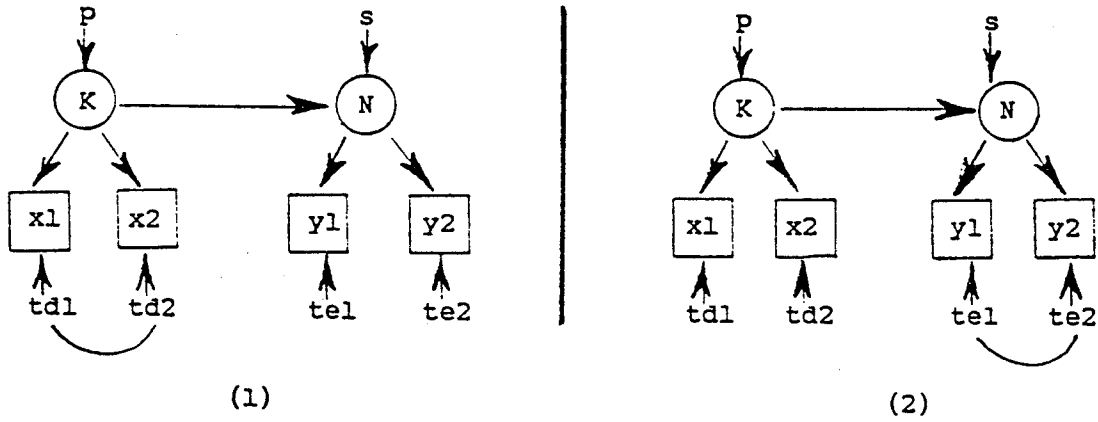


Figure 11-2: Simple Variants Representable in LISREL

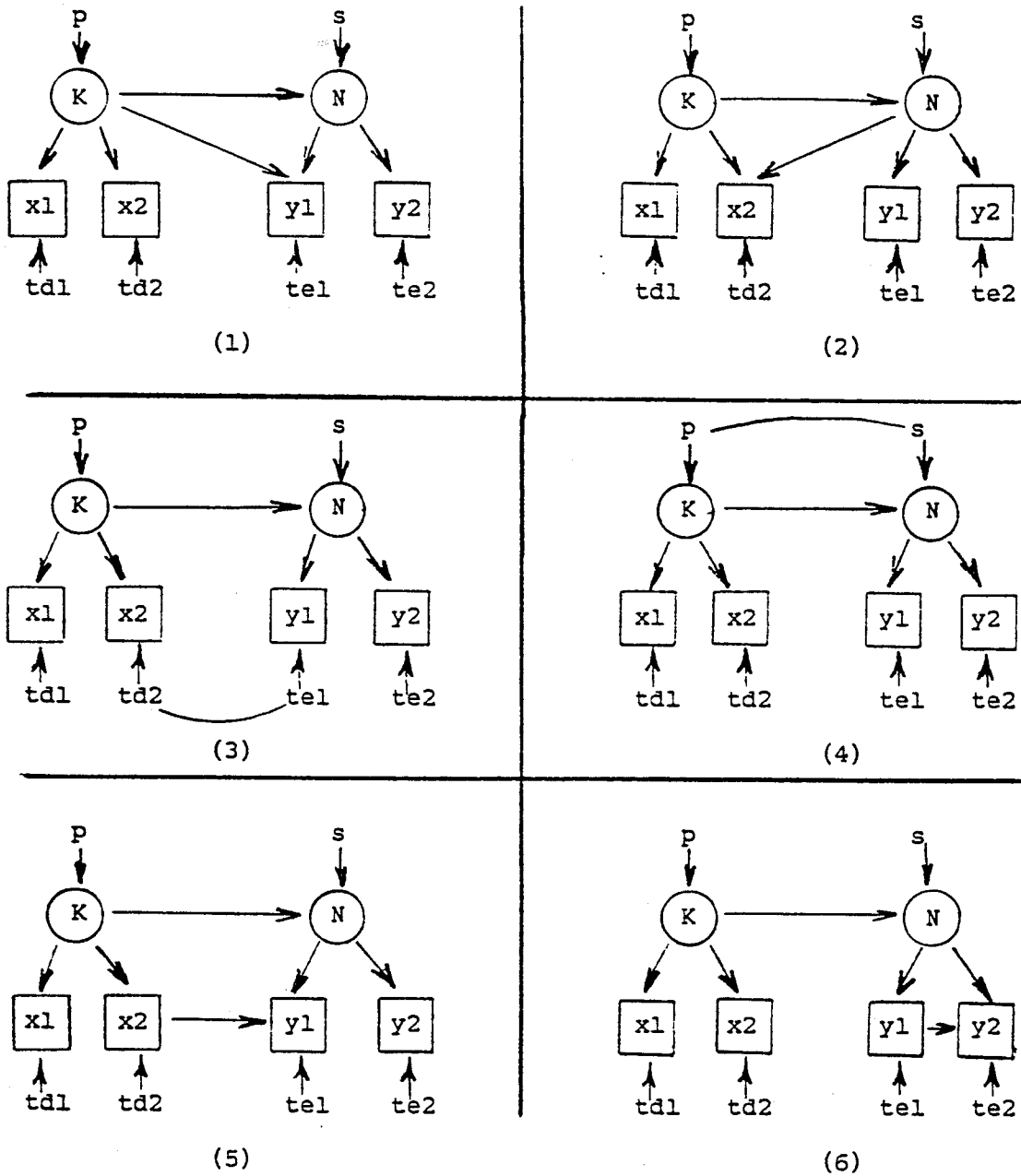


Figure 11-3: Simple Variants Not Representable in LISREL

11.3. OVERCOMING THE RESTRICTIONS

To circumvent LISREL's restrictions, we need only find representations that LISREL will accept and that are appropriate substitutes for the models LISREL will not accept. We

understand a substitution to be appropriate if a model and its substitute are intertranslatable.³⁷ We suggest two ways to do this.

Specifying ksi variables as eta variables

Usually a ksi variable can be specified as an eta variable. You must also change all the x variables that were previously indicators of the ksi variables to y variables that are indicators of the new eta variables, and you must change the error variable types as well. This kind of move will solve the problem of representing variants 1, 2, 3 and 4 of Figure 11-3. We show the solutions as figures 1a, 2a, 3a and 4a in Figure 11-4.

³⁷ That is to say, models I and II are equivalent provided that for every variable of model I there is a definition of that variable in terms of the variables of model II, and for every variable of model II there is a definition in terms of the variables of model I, and the two models, respectively conjoined with their definitions of the terms in the other model, are mathematically equivalent. See Glymour [33, 34] and Glymour and Spirtes [102].

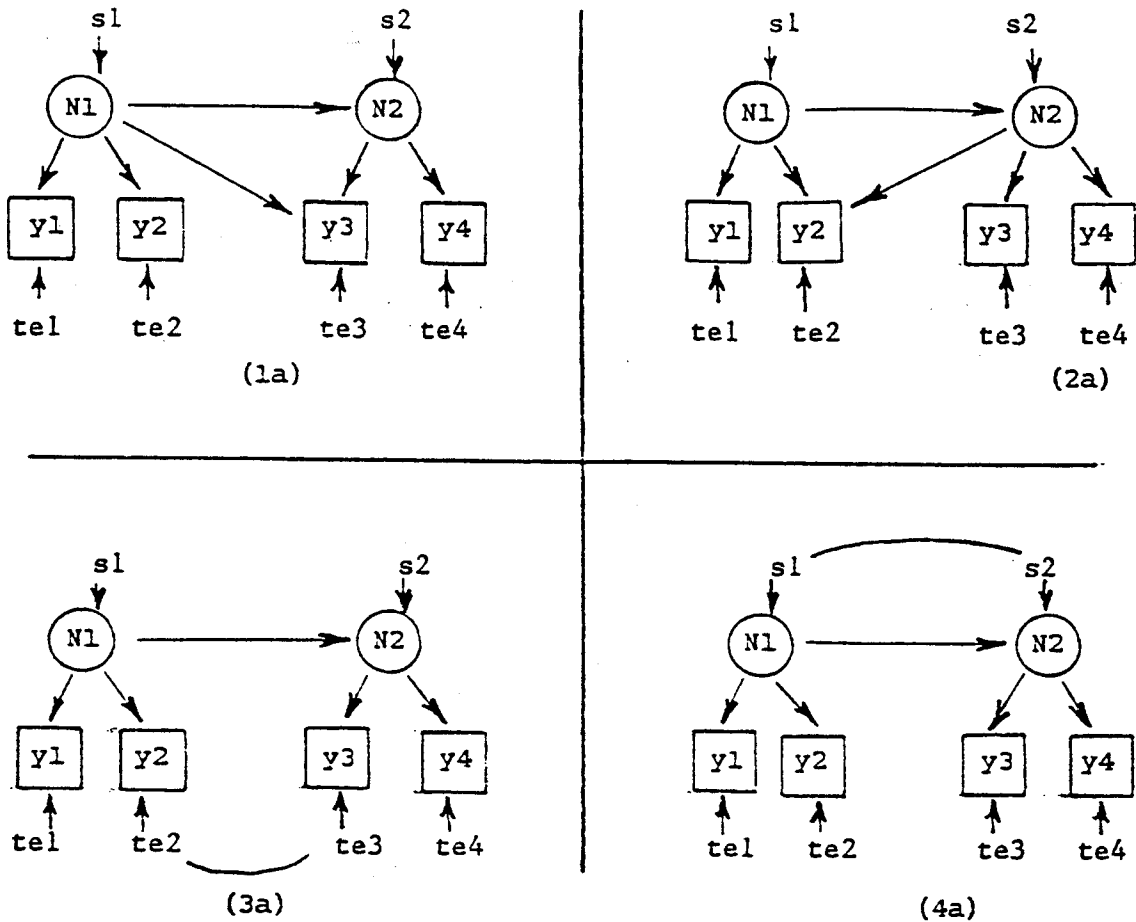


Figure 11-4: Solutions to Variants 1, 2, 3, and 4

This kind of solution is straightforward enough. We are simply changing the LISREL type for certain variables in our model. We are not introducing any new variables, or eliminating any that our original model contained. This technique will work just in case an unacceptable model has an alphabetic variant that LISREL will accept.

Introducing Surrogate Eta Variables

In case a model contains a causal relation between measured variables, e.g., variants 5 and 6 above, the strategy just described is not sufficient. Y variables still cannot cause other y variables. Only LISREL variables of the eta type, and no other, can be represented as both causes and effects. For a LISREL representation of models such as variants 5 and 6, therefore, y variables must be converted into eta variables. This can be done by replacing

a y variable and its error term in the original model with a system of new variables as shown in Figure 11-5. Variables subscripted with an o stand for original, and variables subscripted with an n stand for new.

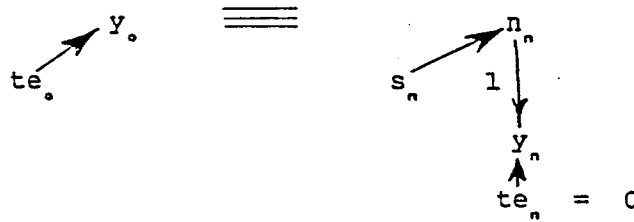


Figure 11-5: Converting a Y Variable Into An Eta Variable and a Y Variable

Notice that s_n occupies the same "function" in the new model that te_o did in the old model. Notice also that the error term for y_n is always fixed at 0. To get the new model from the old model, simply replace every occurrence of y_o with n_n . This does nothing to change the substance of the old model, because in the new model:

$$y_n = 1 * n_n + 0$$

which amounts to saying the two variables are identical. In effect we have created a surrogate variable for the old y which can now enter into causal relationships with other variables of its type. We show the solutions to variants 5 and 6 as solutions 5a and 6a in Figure 11-6.

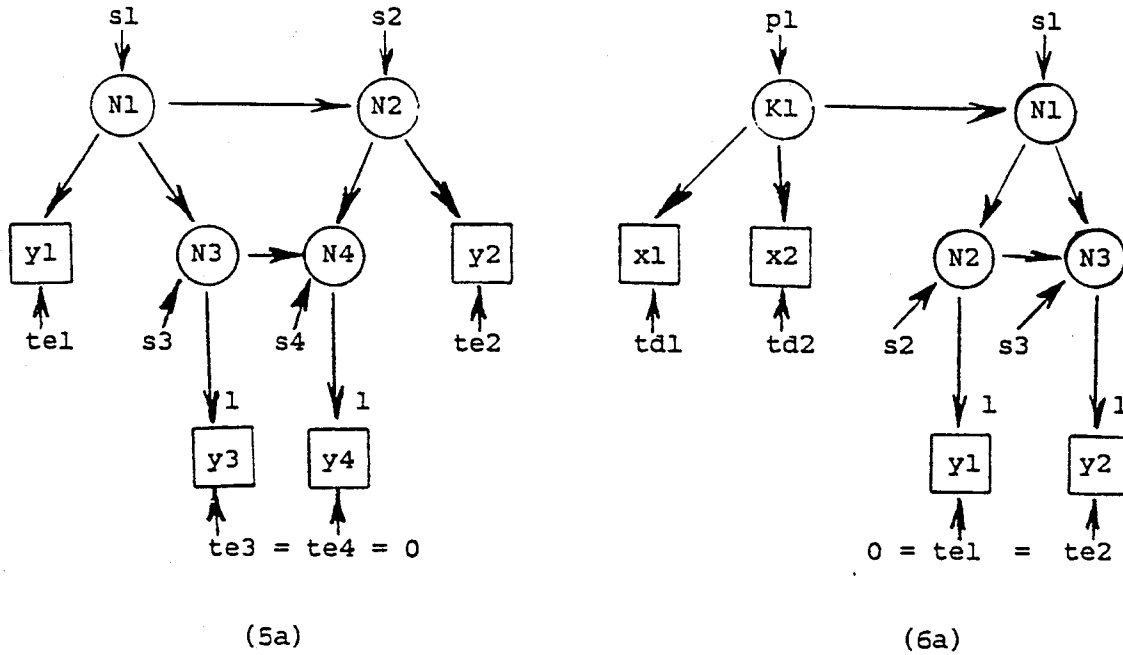
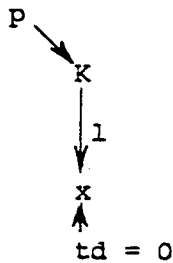


Figure 11-6: Solutions to Variants 5 and 6

Lest users feel uncomfortable with this technique, it should be noted that LISREL has an identical technique already built into its set of options. When a model in LISREL is specified with the option: "FIXEDX" in the model line, this essentially identifies x variables with ksi variables in the same way that we identified y variables with eta variables above. To say "FIXEDX" in LISREL is to create the following situation for each x variable input.



By specifying "FIXEDX" a user can estimate and test a model in which measured variables are exogenous. One is actually creating a surrogate ksi variable for each measured variable; since ksi variables are allowed to be exogenous in the LISREL formalism this

accomplishes precisely the same thing.

11.3.1. An Example: Authoritarian-Conservatism Again

We can illustrate the preceding points by showing the construction of a LISREL input file for a measurement model of the psychological attitude called authoritarian-conservatism discussed in Chapter 8, section 8.4.

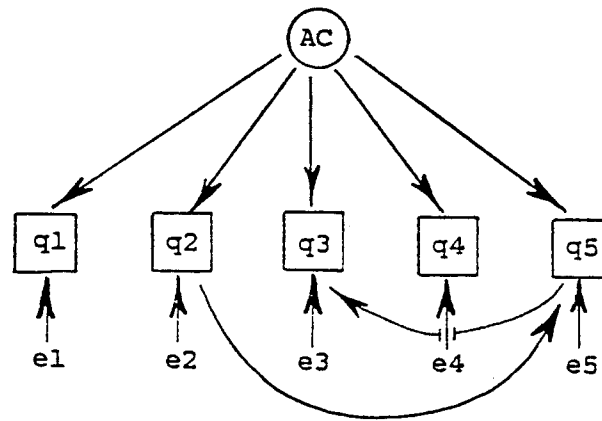


Figure 11-7: Measurement Model to Be Represented in LISREL

To make a LISREL model from this causal model we assume that all errors are uncorrelated. We treat $q1 - q5$ as eta variables, so that we can represent the causal links from $q2$ to $q5$ and $q5$ to $q3$. It would not do to correlate error terms for $q2-q5$ and for $q3-q5$. The model in Figure 11-7 includes a causal chain from $q2$ to $q3$ and thus a source of correlation between $q2$ and $q3$ that is not AC . A model that correlated the error terms for $q2-q5$ and $q3-q5$ would not include a source of correlation between $q2$ and $q3$ that did not stem from AC .

We treat AC as an eta variable that has no indicator. The full LISREL model is shown in Figure 11-8. The variables in parentheses are just to help identify which LISREL variables correspond to which variables in the causal model.

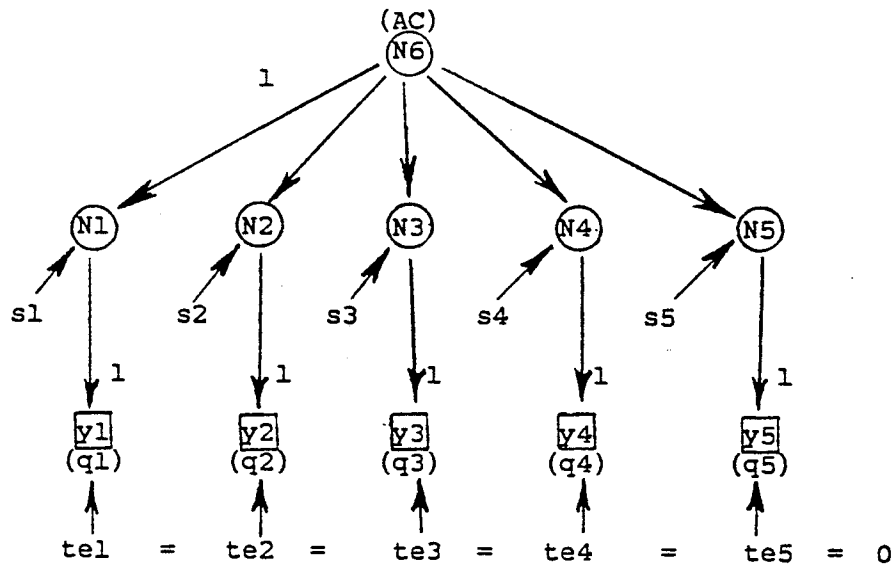


Figure 11-8: Full LISREL Model

We assume the coefficient of the AC \rightarrow q1 path is fixed at unity, and the error terms for all y variables are fixed at 0. One possible LISREL input file for this model is shown beginning on the following page.

ANCHORING: Q2->Q5, Q5->Q3

DA NI=5 NO=3001 MA=KM

LABELS

*

'Q1' 'Q2' 'Q3' 'Q4' 'Q5'

KM SY

*

2.0987

0.7161 2.0286

0.4505 0.3356 1.7258

0.7746 0.5012 0.3525 2.0759

0.4263 0.3421 0.3149 0.3014 1.0658

MO NK=0 NX=0 NE=6 NY=5 GA=ZE LX=ZE TE=ZE TD=ZE PH=ZE PS=DI,FR

PA BE

*

0 0 0 0 0 0

0 0 0 0 0 1

0 0 0 0 1 1

0 0 0 0 0 1

0 1 0 0 0 1

0 0 0 0 0 0

MA BE

*

1 0 0 0 0 -1

0 1 0 0 0 -.5

0 0 1 0 -.3 -.5

0 0 0 1 0 -.5

0 -.3 0 0 1 -.5

0 0 0 0 0 1

MA LY

*

1 0 0 0 0 0

0 1 0 0 0 0

0 0 1 0 0 0

0 0 0 1 0 0

0 0 0 0 1 0

MA PS

*

.5 .5 .5 .5 .5 .5

OU PM TO SE

11.4. EQS

EQS is a recent program authored by Peter Bentler. The input to EQS is much easier than the LISREL formalism, and it is much easier to detect input errors in the EQS program. An EQS input file for the model just considered is pictured below.

```

/TITLE
AC1.DAT: FACTOR MODEL + q2->q5, q5->q3;
/SPEC
VAR=5; CAS=3101; ME=ML;
/LABELS
V1=Q1; V2=Q2; V3=Q3; V4=Q4; V5=Q5; F1=AC;
/EQUATIONS
V1 = F1 + E1;
V2 = .5*F1 + E2;
V3 = .5*F1 + .3*V5 + E3;
V4 = .5*F1 + E4;
V5 = .5*F1 + .3*V2 + E5;
/VARIANCES
E1 TO E5 = .5*;
/MATRIX
2.0987
0.7161 2.0286
0.4505 0.3356 1.7258
0.7746 0.5012 0.3525 2.0759
0.4263 0.3421 0.3149 0.3014 1.0658
/END

```

Figure 11-9: EQS Input File

Changing this input file to create variants of the measurement model pictured in Figure 11-7 is trivial. The problem is in waiting for the output. Examining a serious number of alternative models interactively is prohibitive if one has a personal computer without an 8087 co-processor. To overcome this, EQS can be run repeatedly in batch within a simple ".bat" file.³⁸

When EQS is called from DOS, the personal computer operating system, it loads and then prompts for 3 pieces of information.

1. The input filename.
2. The data filename.

³⁸We are indebted to Eric Wu one of the programmers of EQS, for promptly answering our query with the suggestion we describe here.

3. The output filename

The data filename is optional. In the input file we show above, we have included the covariance matrix as a part of the input file. It is possible to leave the data in a separate file, but we have found it easier to include the data within the input file. If the data is included in the input file, the response to EQS's prompt for a data file is simply a carriage return.

To run EQS in batch you need:

1. A command file that ends in ".bat"
2. A series of files that each contain answers to the 3 prompts EQS gives in a single run.
3. A series of input files.

Suppose we have three variants of the measurement model pictured in Figure 11-7, and suppose the input files for each, with covariance data, are called, respectively:

AC1.dat

AC2.dat

AC3.dat

We would then form three files, called "ln1", "ln2", and "ln3", that contain the answers to EQS' prompts for 3 runs. We show the contents of file "ln1" below.

AC1.dat

AC1.out

Figure 11-10: File "ln1"

"AC1.dat" must occur on the *first* line of the file "ln1". The second line is blank because in this case there is no data file, and a blank line here is the same as a carriage return. The third line tells EQS to send the output to file "AC1.out".

The ".bat" might be called "run.bat", and it would look like this:

```
Eqs <In1  
Eqs <In2  
Eqs <In3
```

Figure 11-11: The File "Run.bat"

To execute the "Run.bat" file, simply type "Run" at the DOS prompt. The first line of the file tells DOS to run EQS, and to get the answers for its prompts from the contents of file "In1". The left angle bracket is a DOS command to change the input stream from the console to a text file. EQS will use the file "AC1.dat" as an input file, it will know there is no separate data file, it will put the output in the file "AC1.out", and when it is finished, DOS will call it again and tell it to now answer its prompts from the contents of file "In2". Any number of models can be run this way.

12. Running TETRAD

12.1. INTRODUCTION

TETRAD is an interactive command driven program which is simple to use. This chapter describes how to use TETRAD. It includes a description of the format of input files, a description of each command, some transcripts of actual TETRAD sessions, and a description of the types of errors that can occur.

12.2. INSTALLING TETRAD

Two version of TETRAD have been provided on your distribution disk. One of these versions (TETRAD) is designed for computers that do not have an 8087 co-processor chip; the other version (TET87) is designed for computers that do have this chip. The two programs act in an identical fashion, except that TET87 is considerably faster than TETRAD.

To run TETRAD you will need to install all of the following files on the same directory.

TETRAD.COM
TETRAD.000
TETRAD.001
HELP.TXT
ERROR.TXT

To run TET87 you will need to install all of the following files on the same directory.

TET87.COM
TET87.000
TET87.001
HELP.TXT
ERROR.TXT

The sample data and graph files may be installed in any directory.

12.3. ENTERING AND EXITING TETRAD

In order to successfully run TETRAD (TET87) your current directory must be the directory in which the TETRAD (TET87) program resides. For example, if TETRAD.COM (TET87.COM) is in directory "\tetrad", then "\tetrad" must be your current directory.

In order to invoke the TETRAD (TET87) simply type "TETRAD" ("TET87") at your monitor prompt.

In the rest of this chapter, we will refer to the TETRAD program, but everything we say will also apply to TET87.

Upon entering TETRAD, a number of messages are printed out. Finally, a ">" is printed. This symbol is a prompt that indicates that TETRAD is waiting for the user to enter a command.

The program is exited by typing "EXIT" when prompted for input.

12.4. GETTING HELP

Online help summarizing the action of each command C is available by entering TETRAD and typing "help C". The list of commands that TETRAD recognizes can be obtained by simply typing "help". A list of each of the commands and the action that it invokes is also provided in this chapter.

12.5. INPUT FILES

The TETRAD program requires both a directed graph and correlation or covariance data for the measured variables in that graph. The file from which TETRAD reads the correlation or covariance data will be called the **covariance file**. The file from which TETRAD reads the directed graph will be called the **graph file**. Using your favorite editor, these files may be created before the TETRAD program is run. Alternatively, they may be created within the TETRAD program using "EDIT", a simple line editor that is part of TETRAD. In this section we explain how to create covariance files and graph files.

12.5.1. Variable Labels

In both the covariance file and in the directed graph, variables must have names. The conventions for naming variables are very simple:

1. Any alpha-numeric character³⁹ can be used in a variable name.

³⁹The set of alpha-numeric characters include all letters, upper and lower case, and all numbers.

2. Each variable name must be 1 or 2 characters long.
3. Names for *measured* variables must begin with a *lower case* letter or a numeric character.
4. Names for *latent*, or unmeasured variables must begin with an *upper case* letter.
5. One and the same variable must have the same name in the covariance file and in the directed graph.
6. Upper case letters are distinct symbols from lower case letters.

12.5.2. Covariance Files

A TETRAD covariance file is a normal text file. With the IBM PC operating system, MS DOS, as many as eight letters can occur in the filename and three more in an extension. Other operating systems may have different conventions. The extension is not necessary although it can be useful. Examples of data filenames are:

```
Datal  
Datal.pas  
ladata  
Theory1.dat
```

The first non-blank line of a covariance file must be an integer which represents the sample size. It occurs on a line by itself.

After the sample size, the only information in a covariance file is a correlation or covariance for each pair of measured variables. The user may enter either correlational data or covariance data in the covariance file; TETRAD converts all covariances into correlations. In what follows we mean "correlation or covariance" whenever we say just "covariance". By a "number" we always mean a real number or integer. There are two distinct ways of representing the covariance data in a file.

12.5.2.1. Covariances on Each Line

In the first way of representing covariances, a separate line in the covariance file must be created for each covariance. Each line contains, in order:

1. One correlation, entered as a number.
2. The two variable names which correspond to the covariance, separated by at least one space from the covariance number and from each other.

3. A carriage return.

No latent variables may occur in the covariance file. Every pair of measured variables appearing in the covariance file must have exactly one number associated with it in the covariance file. (Each variable must also be paired with itself; that is for each variable x , there must be an entry for the pair $x\ x$.)

An example is:

```
.576 x1 x2
```

The order of the variable names in a line does *not* make a difference. The covariance in the example could just as well be entered as:

```
.576 x2 x1
```

Note, however, that it would be illegal to enter both of the above lines in a single covariance file, since that would assign more than one (albeit equal) covariance to a single pair of variables.

The covariances do not have to be entered in any particular order. The sequences in Figure 12-1 (which are all *fragments* of covariance files) are all the same so far as the TETRAD program is concerned.

<pre>.24 x1 x2 .45 x1 x3 .86 x2 x3</pre>	<pre>.45 x1 x3 .86 x2 x3 .24 x1 x2</pre>	<pre>.86 x3 x2 .24 x1 x2 .45 x3 x1</pre>
(1)	(2)	(3)

Figure 12-1: Equivalent Input Sequences

The fragments of covariance files shown in Figure 12-2 are illegal and will not be read correctly by the program.

<pre>.24x1 x2 .45 x1 x3 .86 x2 x3 (There is no space between) (the correlation and the) (first variable name.)</pre>	<pre>Correlations .24 x1 x2 .45 x1 x3 .86 x2 x3 (Titles are not allowed.)</pre>
<pre>x1 x2 .24 x1 x3 .45 x2 x3 .86 (The correlation must come first.)</pre>	<pre>.24 x1 x2 .45 x1 x3 .86 x2 x3 (Only one correlation is) (allowed per line.)</pre>

Figure 12-2: Illegal Covariance Files

To illustrate these points, suppose we have constructed a simple measurement model for socio-economic status, with four indicators: father's education, father's occupation, mother's education, and family income. Our abbreviation scheme is as follows.

```
socio-economic status = SE
father's education    = fe
father's occupation   = fo
mother's education    = me
family income         = in
```

We show our hypothetical model in Figure 12-3.

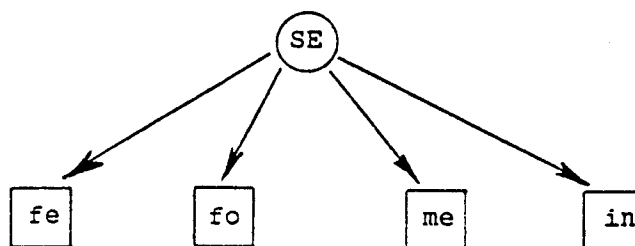


Figure 12-3:

A graph file for the above example would be:

```
SE fe  
SE fo  
SE me  
SE in
```

Figure 12-7: Example of a Graph File

12.6. OUTPUT FILES

The user may specify the name of a file where output is to be sent by using the "OUTPUT" command. The "OUTPUT" command will prompt the user for the file name. If the output file has not been specified, then any command producing output will prompt the user for it. After the output file has been specified, all subsequent output will be sent to that file until the user issues another "OUTPUT" command that specifically changes the output file.

12.7. VIEW AND EDIT

Two small subsystems with their own set of commands have been incorporated into TETRAD. "VIEW" is for viewing output files, and "EDIT" is for creating or modifying input files. Since each of these commands reads files into memory, using them on large files will use large amounts of memory. TETRAD will produce an error message if the file is too large to fit into memory. Each of these subsystems is intended for use on files which have maximum line lengths of 80 characters.

12.8. THE RUN COMMAND AND MENUS

The easiest way to use TETRAD is through the "RUN" command. Each output command, except the "RUN" command, is used to produce output about one specific feature of a model. The purpose of the "RUN" command is to allow the user to use a single command to produce output about a variety of features of a model.

The "RUN" command is used in conjunction with the "SETOUTPUT" command. The "SETOUTPUT" command provides the user with a menu of possible outputs that can be produced by the "RUN" command. Any combination of desired outputs can be produced by choosing the desired outputs from the menu of the "SETOUTPUT" command, and then issuing the "RUN" command. It should be emphasized that the outputs selected from the

"SETOUTPUT" command do not themselves produce any output, nor do they have any effects on any command except the "RUN" command. An example of the use of the run command is shown below. Note that "RUN" prompts for any information that it needs that has not already been specified; if the information had already been specified, then the information would not have been prompted for. Further details about "RUN" and "SETOUTPUT" are in section .

The "SETPARAMETER" command prompts the user for all of the program parameters. Unlike the "SETOUTPUT" command, this command affects all of the output output commands.

While many of the other commands make use of TETRAD more convenient, most of the features of TETRAD can be invoked using just these three commands, together with the "EXIT" command.

The following is a transcript of an actual TETRAD run. Input typed in by the user has been placed in boldface. User input is always followed by a carriage return which is generally not shown. However, on lines where the user enters only a carriage return, this has been symbolized by <CR>. Any italic text between a '{' and a '}' is a comment that has been added to the transcript to explain the commands. All of the examples in this chapter will assume that the covariance file used was d1.dat and the graph file used was d1graph (both shown below).

D1.dat has the following contents.

```
500
1.0 x1 x1
1.0 x2 x2
1.0 x3 x3
1.0 x4 x4
1.0 x5 x5
1.0 x6 x6
1.0 x7 x7
1.0 x8 x8
1.0 x9 x9
0.562 x1 x2
0.658 x1 x3
0.471 x2 x3
0.750 x1 x4
0.524 x2 x4
0.866 x3 x4
0.774 x1 x5
0.550 x2 x5
0.790 x3 x5
0.969 x4 x5
0.783 x1 x6
0.563 x2 x6
0.674 x3 x6
```

0.880 x4 x6
 0.949 x5 x6
 0.779 x1 x7
 0.565 x2 x7
 0.674 x3 x7
 0.875 x4 x7
 0.942 x5 x7
 0.989 x6 x7
 0.760 x1 x8
 0.555 x2 x8
 0.655 x3 x8
 0.855 x4 x8
 0.920 x5 x8
 0.956 x6 x8
 0.957 x7 x8
 0.787 x1 x9
 0.566 x2 x9
 0.674 x3 x9
 0.878 x4 x9
 0.946 x5 x9
 0.993 x6 x9
 0.990 x7 x9
 0.958 x8 x9

D1graph has the following contents.

T1 T2
 T2 T3
 T1 T3
 T1 x1
 T1 x2
 T1 x3
 T2 x4
 T2 x5
 T2 x6
 T3 x7
 T3 x8
 T3 x9

In the following TETRAD session, the user sends output about the suggested sets and the tetrad equations to a file called firstout. The transcript is as follows.

```

C:\tetrad          { The program is invoked. }
COPYRIGHT (C) 1984 by R. Scheines, C. Glymour, K. Kelly, and P. Spirtes
All Rights Reserved
  
```

For help, type "help"

```

{ The user plans to use the "RUN" command. The current values of
  all of the output options is printed. Then the user is
  prompted for a list of numbers of output options which he wants to
  toggle. In this case the user wants to see all of the suggested
  sets, and the implied tetrad equations. }
  
```

>setout

Tetrad Equations

Vanishing Partial

1: Suggested sets FALSE

Statistical Info On:

2: All tetrad FALSE
3: Implied tetrad FALSE
4: Held tetrad FALSE

Statistical Info On:

5: All partial FALSE
6: Implied partial FALSE
7: Held partial FALSE

Rttr Chart Info for:

9: Measured - Measured Edges: FALSE
10: Latent - Latent Edges: FALSE
11: Latent - Measured Edges: FALSE
12: Measured - Latent Edges: FALSE

8: Rtpr Chart: TRUE

Enter the numbers of options that you want to toggle: 1 2

{ The output options with the new values are now automatically printed out. }

Tetrad Equations

Vanishing Partial

1: Suggested sets TRUE

Statistical Info On:

2: All tetrad TRUE
3: Implied tetrad FALSE
4: Held tetrad FALSE

Statistical Info On:

5: All partial FALSE
6: Implied partial FALSE
7: Held partial FALSE

Rttr Chart Info for:

9: Measured - Measured Edges: FALSE
10: Latent - Latent Edges: FALSE
11: Latent - Measured Edges: FALSE
12: Measured - Latent Edges: FALSE

8: Rtpr Chart: FALSE

{ No output has been created so far. The "RUN" command is now invoked to produce output about the suggested sets and the tetrad equations. Since some information needed to run the program has not been entered yet, the command prompts for the covariance file, the graph file, the significance level, and the name of the file where output is to be sent. In this case, the output is sent to a file called firstout. }

>ru

Covariance file: d1.dat

Graph file: d1graph

Edge: T1 --> T2
Edge: T2 --> T3
Edge: T1 --> T3
Edge: T1 --> x1
Edge: T1 --> x2
Edge: T1 --> x3
Edge: T2 --> x4
Edge: T2 --> x5
Edge: T2 --> x6
Edge: T3 --> x7

```
Edge: T3 --> x8
Edge: T3 --> x9
```

```
{ The default value of the significance level is 0.05. }
Significance level: [0.05000]: .4
```

```
Output file: firstout
```

```
> exit
```

12.9. USER INTERRUPTS

Some of the calculations that TETRAD performs can be quite time-consuming, especially if the model is a large one. For that reason, the user has been provided with a way to interrupt some calculations and return (almost) immediately to the TETRAD command prompt. The user can interrupt a calculation by holding down the control key and *simultaneously* pressing the "g" key. This will only interrupt the calculation of the Rtrr chart, and the calculation of tetrad or partial equations for a model.

12.10. ERRORS

Two kinds of errors can be made when using TETRAD. **Syntax errors** are "grammatical" errors in the input; they include such things as improper variable names or illegal numbers.

Semantic errors in TETRAD are errors in the "meaning" of the input; they include such things as putting in the wrong covariance between two variables or specifying an edge between two variables which should not be connected. Naturally, TETRAD cannot detect all semantic errors; it has no way of knowing whether the covariance between two variables should be 0.01 or 1000. However, TETRAD does detect some semantic errors.

12.10.1. The Form of Error Messages

An error message in TETRAD consists of either four or five parts, depending upon whether the error occurs in input that is entered from a terminal, or in data read in from a file. These parts are:

- An explanation of the action TETRAD is going to take.
- If the error is from a file, the name of the file and the line number in which the error occurred.

- The line on which the error occurred.
- A pointer showing the place in the line at which the error occurred.
- An explanation of the error.

For example, suppose a graph file contained the following line: "T1 x%". TETRAD would have the following response.

```
Warning: Input being ignored.
T1 x%
  ^
  Non-alphanumeric character in tail name.
```

12.10.2. Actions

When TETRAD discovers an error in a line of input, there are three different kinds of actions that the program may take.

- TETRAD may simply ignore a line of input, and continue on. In this case, the line of input is simply treated as if it did not exist. Upon taking this action TETRAD writes "Warning: Input being ignored."
- TETRAD may ignore part of a line of input. In this case, the beginning of the line is read, and the end of the line is simply treated as if it did not exist. Upon taking this action TETRAD writes "Warning: " at the beginning of the message. At the end of the message it points to a character in the input line; all characters after that are ignored. TETRAD then writes "Rest of line ignored."
- If the error prevents TETRAD from continuing to process its data, TETRAD will return to the command prompt without creating any output. The next time a command creating output is issued, TETRAD will prompt the user to re-enter the data. For example, if covariances between all of the observable variables are not contained in the covariance file, TETRAD will prompt the user for a new covariance file. In this case TETRAD writes "Fatal error: " at the beginning of the error message.

12.10.3. Agreement between Graph and Covariance Files

TETRAD receives input of two different kinds from two different sources. A covariance file specifies the correlations or covariances between various observable variables, and a graph file contains a directed graph. The covariance file should contain covariances for every pair of observable variables that appear in the graph.

It is possible that the input file might not specify covariances for every pair of observable variables in the graph. Alternatively, the covariance file might not specify covariances for every pair of observable variables occurring in the data file itself. In either case the error is fatal. TETRAD will not produce any output, and will give control to the part of the program which asks the user if he or she would like to enter a new graph.

Sometimes one may wish to run TETRAD using a directed graph that contains some but not all of the measured variables in the covariance file. And sometimes one may intend to enter a directed graph that contains all of the measured variables in the covariance file, but inadvertently omit some of those variables. TETRAD cannot intuit the user's intentions, but the program does recognize the ambiguity of the situation. If the covariance file contains covariances between pairs of observable variables that do not occur in the graph, TETRAD simply issues a warning, and continues normally.

12.10.4. Maximum Number of Variables and Edges

TETRAD sets a limit to the maximum number of variables that can appear in an input file and graph. The exact value of this variable depends upon the kind of computer being used. Currently, for the IBM PC and compatibles, the maximum number of variables is 23 (9 for the demonstration disk accompanying this book). If the maximum number of variables is exceeded the following error message will appear. If there are variables in your covariance file that do not appear in the graph, then removing all occurrences of these variables from the input file may reduce the total number of variables to below the maximum number allowed.

One distinctive error message that TETRAD can issue concerns the maximum number of edges that may be entered in a graph. When this total is reached, TETRAD will stop prompting for more edges and print a warning.

12.10.5. Memory

Memory errors occur when the computer's random access memory capacity is exceeded. Such errors may occur because:

1. The model you are trying to run is too big for TETRAD.
2. The model you are trying to run is too big for your computer's memory.

3. The significance level is too large for the model you are analyzing.⁴⁰
4. Too many units of suggested sets have been requested.

12.10.6. Semantic Errors TETRAD Will Not Recognize

We list below some mistakes that are likely to cause TETRAD to give poor results. They are especially difficult to locate because syntactically there is nothing to cause an error message from TETRAD. Everything looks fine, but when TETRAD's recommendations are subjected to other methods of evaluating linear models (e.g. LISREL's chi-square calculator), the results are disastrous.

1. The *Data File* is *faulty*.

There might be a typo in the correlations (e.g. a number was entered incorrectly), or there might be incorrect variable names bound to a correlation.

2. The *Graph* is *faulty*.

- a. The graph for which suggested trek additions were requested is not skeletal and has a measured variable directly connected to two or more latent variables.
- b. The wrong graph was entered.

3. The *General Assumptions* were *faulty*.

- a. The initial skeleton is a poor model.
- b. The assumptions of structural equation models (See Chapter 2, section 6) do not hold.
- c. The sample is inadequate or unrepresentative.

12.11. TETRAD COMMANDS

TETRAD is not sensitive to the case of the characters that a user types in (except for variable names). Also, TETRAD recognizes any non-ambiguous substring of a command. For example, if a user enters "ru", TETRAD will recognize the "RUN" command. However, if

⁴⁰ TETRAD is forced to do more computation on the same model and data set for larger values of significance level. When significance level is increased fewer equations "hold" in the data, consequently fewer edges are prohibited from being suggested as additions. The computational demands go up when the number of admissible candidates to a suggested set of treks go up.

a user enters "r", then TETRAD will print an error message pointing out that "r" is ambiguous (because there is a "RUN" command and a "RTTR" command).

12.11.1. Types of Commands

TETRAD commands may be divided into three broad categories.

- Output-producing commands cause information about a model to be written to a file. An example of an output command is "PARTIAL", which outputs information about which partial equations are implied by a model, and which partial equations hold in the population. (Note that the "OUTPUT" command does not itself produce output, and so is not an output-producing command).
- Parameter commands set various parameters (such as the significance level) which affect the output produced by the output commands. For example the "SIGNIFICANCE" command allows the user to reset the significance level which determines whether a particular equation is said to hold in the data or not. Such a command affects the output produced by output commands such as "PARTIAL", but does not itself produce any output. To invoke a given parameter command, simply type the name of the command; the command will then prompt for the new value of the parameter.
- Finally there are miscellaneous commands that are helpful in running the program, but do not produce or effect the output of the program. For example, the "VIEW" command allows the user to view output files produced by TETRAD.

12.11.2. Switches

Each output-producing command has a (possibly empty) set of compatible switches. Adding a switch to a command limits the output of the command to some subset of the normal output. Switches must be entered on the same line as the command itself. For example, the "RTTR" command has four compatible switches, "MM", "ML", "LM", and "LL". If the user types "Rttr", then the entire Rttr chart is output. If the user types "Rttr mm", then only that part of the Rttr chart involving an edge between two measured variables is output.

It is possible to enter more than one switch on a command line. Any subset of the switches that are compatible with a given command may be entered in any combination and in any order. The effect of entering more than one switch on a command line is cumulative; that is, the output produced by a combination of switches combines all of the output that would have been produced by each switch individually.

For example, "rttr mm ml" is equivalent to "rttr ml mm". The effect of both of these

command is to output that part of the rtrr chart involving an edge between two measured variables, or an edge between a measured variable and a latent variable.

12.11.3. Prompting

Each output-producing command has certain required information that it needs to produce its output. For example, it is impossible to calculate which tetrad equations are implied unless a model has been specified. For that reason, if an output command is entered and some of the necessary information is missing, the command will prompt the user for the missing information. If the prompt is followed by a value between square brackets ('[', ']'), that value is the default value. Simply hitting a carriage return at that point enters the default value.

12.12. RUNNING TETRAD IN BATCH MODE

Versions of MS-DOS labelled 2.1 and higher are capable of redirecting input and output. To get input from a file named "A" instead of from the terminal simply type "tetrad < A". "A" should contain all of the commands to TETRAD that one would normally enter from the terminal. Below is an example of a file which produces a TETRAD session identical to the example in section , but in which the input is from a file instead of the terminal.

```
setout
1 2
ru
dl.dat
dlgraph
.4
firstout
exit
```

Note that the file includes not only commands such as "SETOUTPUT" and "RUN", but also answers to prompts that TETRAD issues.

To run TETRAD in batch mode simply create a batch file with a ".bat" extension. If input is to come from file "A", simply put in the batch file the command "tetrad < A". Then invoke the batch file by typing its name (without the ".bat" extension). For example, a file "tet.bat" could be created. The contents of tet.bat would be the line "tetrad < A". To invoke tet.bat simply write "tet".

12.13. COMMAND SUMMARY

12.13.1. Introduction

This section contains a brief list of all of the TETRAD commands together with a brief summary of what each does. Following the brief list of commands, each command is described in more detail. This section does i(not) explain what the output means. It uses a number of concepts explained in chapters 7 and (m-meth).

12.13.2. List of Commands

Output-producing commands: All send output to current output file. Each will prompt for any parameter values that are needed but that have not been set by the user.

Compare	Detailed effects of adding edge to graph
Partial	Partial equations
Rttr	Effect on tetrad equations of adding edges to graph
Rtpr	Effect on partial equations of adding edges to graph
Run	Various outputs depending on setoutput
Standard	Summary information
Suggested	Sets of edges which will improve graph
Tetrad	Tetrad equations

Parameter Commands: Change parameter values

Changegraph	Changes current graph
Input	Covariance and graph files
Output	Output file
Pi	Sets parameters of base model for pi calculation
Setparameters	Prompts for information about all parameters
Significance	Resets the significance level
Units	Number different significance levels for suggested sets

Miscellaneous Commands

Edit	Line editor for input files
Exit	Exits program
Help	Help on each command
Setoutput	Menu to set various parameters for run command
Status	Current value
View	Views output files a screen at a time

Parameters: The following are not TETRAD commands. They are user settable parameters that affect the output of output-producing commands, and are set by parameter commands.

- Covariances
- Current graph
- Output file
- Significance level
- Units

Number of edges in base model
 TTR of base model
 Number of latent variables of indegree 0 in base model

12.13.3. Command Summary

In this section a brief description of each command is given. The format of each description is the same.

- The name of the command.
- A brief description of the action taken by the command.
- If it is an output-producing command, a description of the switches that can be used with the command. This includes:
 - The default switch (i.e. what happens if no switch is specified with the command).
 - The name of a switch.
 - The effect of using that switch.
- The type (output-producing, parameter, or miscellaneous command).
- A list of the parameters that affect the output of the command (in the case of output-producing commands) or a list of the parameters that the command affects (in the case of parameter commands). This section is left out in the miscellaneous commands.
- If it is an output-producing command, a list of other commands that interact with the given command.

CHANGEGRAPH

Changegraph prompts for a series of edges to be added to or deleted from the current graph. First it prints out the current graph. Then it prompts for edges to delete. If there are edges to delete, the user should enter one of these edges, and press carriage return. The format of an edge is exactly the same as that used in graph files, that is, a variable name followed by at least one space, followed by a second variable name. The command will continue to prompt for edges to delete until the user enters a carriage return at the prompt. This is taken as a signal to stop prompting for further edges to delete. Then the command prompts for edges to add in the same fashion. Finally it prints out the graph with the changes the user has made. While it is possible to delete *edges* it is not possible to delete *vertices*; a vertex is present even after all of the edges containing it have been deleted.

The following is an example of the use of changegraph to add the edge x1 x2 to the current graph.

```
>change
Edges to be deleted.
```

Edge: <CR>

Edges to be added.

Edge: x1 x2

Edge: x1 --> x2

Edge: <CR>

Type

Parameter

Affects parameter

Current graph

COMPARE

Compare prompts the user for an edge to add to the current graph. Compare prints out all of the equations which were implied in the original graph but are not implied by the graph with the added edge. (Note that it is not possible for an equation to be implied in the extended graph but not in the original graph). For each equation it prints out whether or not it is implied by the *extended graph*, whether or not it holds in the data, its residual, and its p value. At the end it specifies the Rtrr, D(I-H) and D(H-I) values for that edge. Below is an example of the use of the compare command and the output it produces. (Remember we are assuming the covariance file is d1.dat and the graph file is d1graph).

```
>compare
Edge: x1 x2
>
```

The relevant part of the output produced by this command was:

Edge added: x1 x2 Tetrad Equation	Residual	Impld.	Held	P(diff.)
x1 x2, x4 x3 = x1 x4, x2 x3	0.1334			0.0000
x1 x2, x3 x4 = x1 x3, x2 x4	0.1419			0.0000
x1 x2, x5 x3 = x1 x5, x2 x3	0.0794			0.0018
x1 x2, x3 x5 = x1 x3, x2 x5	0.0821			0.0000
x1 x2, x6 x3 = x1 x6, x2 x3	0.0100		y	0.6631
x1 x2, x3 x6 = x1 x3, x2 x6	0.0083		y	0.6574
x1 x2, x7 x3 = x1 x7, x2 x3	0.0119		y	0.5992
x1 x2, x3 x7 = x1 x3, x2 x7	0.0070		y	0.7127
x1 x2, x8 x3 = x1 x8, x2 x3	0.0102		y	0.6519

x1 x2, x3 x8 = x1 x3, x2 x8	0.0029	y	0.8816
x1 x2, x9 x3 = x1 x9, x2 x3	0.0081	y	0.7237
x1 x2, x3 x9 = x1 x3, x2 x9	0.0064	y	0.7326
Rttr: 0.502 D(I-H): 4 D(H-I): 8			

Each of these equations was implied by the original graph, but is not implied when the edge x1 x2 is added to the graph.

Switches

Default switch: tetrad

Common: Causes the two variables entered by the user to be interpreted as specifying a correlated error between those variables.

Tetrad: Print out tetrad equations.

Partial: Print out partial equations.

Type

Output-producing

Parameters that affect

Covariances, Current Graph, Significance

See also

Input, Changegraph, Significance

EDIT

Edit is a simple line editor. It prompts for a filename to edit. If the file currently exists, its contents are read in to a buffer. Edit always has a current line, which is where commands that change the file act. Initially, the current line is set to the first line of the file. 'u' and 'd' move the current line up and down respectively. When in append mode, the user may enter a series of lines which will be entered into the file just after the current line. Append mode is ended when the user types a '.' as the first character of a line. Insert mode is similar except that the text is entered just prior to the current line. It recognizes the following commands. The changes made are not written to the file until the 'w' command is issued.

Change to: T3 x7

w { *The user save the file.* }
 File name [d1.dat]: d1.dat
 e { *The user exits.* }

Type

miscellaneous

EXIT

Causes the program to exit.

Type

miscellaneous

HELP

Simply typing help gives a brief summary of each TETRAD command. For a given command C, typing "help C" give a more extensive description of the command. Unlike the command interpreter, "HELP" does not accept abbreviations of commands.

Help prints its information a screenful at a time, and then pauses until the user presses another key. Pressing "e" or "E" exits help and returns to the command prompt. Pressing any other key except <CR> prints the next screenful of information.

The information on each command includes what affect issuing the command has, what parameters it is affected by or affects, and how each switch modifies the action of the command.

Type

Miscellaneous

INPUT

Input first prompts the user to enter the name of a covariance file, and then prompts the user to enter the name of a graph file. If a file does not exist, the user will be re-prompted for a file name, until the user enters the name of a file that does exist. The default value for each file is the current file.

Type

Parameter

Affects parameters

Covariances, Current Graph

OUTPUT

Output changes the file where output is sent to <filename>. If the file already exists, then the user will be asked if he wishes to append further output to the file. If the user answers no to this question, then the user will be asked if he wishes to overwrite this file. If the answer to this question is also no, then the user will be asked to enter another file name. If the user wishes to send output to the terminal, then he or she should enter the name "con:" in response to the prompt.

Type

Parameter

Affects parameter

Output file

PARTIAL

Partial prints out information about a set of partial equations. The membership of the set is determined by the switches. For each member of the set partial prints out whether it is implied, whether it holds at the given significance level, and the probability of obtaining the observed partial residual given that the residual in the population is 0.

Switches

Default switch: All

All: Information about every partial equations among each triple of distinct observable variables. If this switch is on, the other switches are ignored.

Held: Information about every partial equation among each triple of distinct observable variables that holds at the given significance level.

Implied: Information about every implied partial equation among triples of distinct observable variables.

Type

Output-producing

Parameters that affect

Covariances, Graph File, Significance

See also

Input, Changegraph, Significance

Parameters that affect

Covariances, Current graph, Significance, Number of edges in base model, TTR of base model, Number of latent variables of indegree 0 in base model

See also

Input, Changegraph, Significance, Pi

PI

Pi is a command used to set the parameters of the base model. In order to calculate the pi value of a model, three things must be known about the base model:

- The number of edges.
- The TTR.
- The number of latent variables with 0 indegree.

"PI" first asks if the user wishes to use a model with one latent variable connected to every other variable by a single edge as the base model. (Henceforth, such a model will be called a **star model**. If the user answers "y" or "Y", then TETRAD will automatically set the parameter values of the base model to equal those of the star model. The covariances among variables are those specified in the covariance file. The variables in the star model are those appearing in the graph file. If these files have not been specified then

TETRAD will prompt for them. The advantage of using a star model as a base model is that it makes the pi value of any model non-negative, since it implies all possible tetrad equations. If the "PI" command is never issued, the star model is what TETRAD uses as a base model.

If the user answers no to the first question, then he or she is asked if the current model should be the base model. If the user answers "y" or "Y", then TETRAD will automatically set the parameter values of the base model to equal those of the current model. Once again, the current model is based upon the current graph and covariance files; if these have not been specified then TETRAD will prompt for them.

If the user answers no to the second question, then TETRAD will prompt the user to enter the number of edges, TTR, and number of latent variables of zero indegree for the base model.

Once a "PI" command has been issued, the base model will remain unchanged until the next "PI" command is issued. (Upon entering TETRAD, if no "PI" command is issued, the parameters of the base model are based upon the star model derived from the first covariance and graph files entered.)

Type

Parameter

Affects parameters

Number of edges in base model, TTR of base model,
Number of latent variables of indegree 0 in base model

RTPR

For each edge in a given set, RtpR prints out what effect adding that edge to the current graph will have (as long as the RtpR is greater than 0). The set of edges includes all edges that are not currently in graph, and correlated errors between each pair of variables that do not have an edge in either direction between them. The information printed includes how much the TPR of the current graph will be reduced, how much (PI-PH) will be reduced, and how much (PH-PI) will be increased. The names given to these quantities are:

I(PH-PI): The increase in the number of equations that hold, but aren't implied.

D(PI-PH): The decrease in the number of equations that are implied, but don't hold.

RtpR: The reduction in the total tetrad residual.

Switches

none

Type

Output-producing

Parameters that affect

Covariances, Current graph, Significance

See also

Input, Changegraph, Significance

RTTR

For each edge in a given set, Rtrr prints out what effect adding that edge to the current graph will have (as long as the Rtrr of the added edge is greater than 0). The information printed includes how much the TTR of the current graph will be reduced, how much (I-H) will be reduced, how much (H-I) will be increased and the pi value of the modified model. The names given to these quantities are:

- I(H-I): The increase in the number of equations that hold, but aren't implied.
- D(I-H): The decrease in the number of equations that are implied, but don't hold.
- Rtrr: The reduction in the total tetrad residual.
- Pi: The pi value of the model.

Switches

Default Switch: All

- All: Each edge that is not already present in the graph, and each correlated error between pairs of observable variables that have no edge in either direction between them. If this switch is on, the other switches are ignored.
- LI: Each latent-latent edge that is not already in the graph.
- Lm: Each latent-measured edge that is not already in the graph.

- Ml:** Each measured-latent edge that is not already in the graph.
- Mm:** Each measure-measured edge that is not already in the graph, and each correlated error between pairs of measured variables that have no edge in either direction between them.

Type

Output-producing

RUN

Run combines the output of all of the other output-producing commands in a single command. The exact output of Run depends upon the options set via the setoutput command. However, run always produces the output that is produced by the standard command.

Switches

none

Type

Output-producing

Parameters that affect

Covariances, Current graph, Units, Significance
 Number of edges in base model, TTR of base model,
 Number of latent variables of indegree 0 in base model

See also

Input, Changegraph, Setoutput, Units, Significance, Pi

SETOUTPUT

Setoutput determines what the output of the run command will be. It prints the following menu.

<u>Tetrad Equations</u>	<u>Vanishing Partial</u>
1: Suggested sets FALSE	
Statistical Info On:	Statistical Info On:
2: All tetrad FALSE	5: All partial FALSE
3: Implied tetrad FALSE	6: Implied partial FALSE
4: Held tetrad FALSE	7: Held partial FALSE
Rttr Chart Info for:	8: Rtpr Chart: FALSE
9: Measured - Measured Edges: FALSE	
10: Latent - Latent Edges: FALSE	
11: Latent - Measured Edges: FALSE	
12: Measured - Latent Edges: FALSE	

The boolean values next to each option indicate whether it is currently on or off. The default value for each option is false. In order to toggle the value of any set of options, simply type the number of each option at the prompt. After the numbers of the options have been typed in, the menu is printed again so that the user may see the effect that his input has had on the selected options.

The output produced by the run command if a given option is on is equivalent to the output of one of the other output-producing commands. These equivalences are described below.

- 1: Suggested sets: suggestedsets
- 2: All tetrad: tetrad all
- 3: Implied tetrad: tetrad implied
- 4: Held tetrad: tetrad held
- 5: All partial: partial all
- 6: Implied partial: partial implied
- 7: Held partial: partial held
- 8: Rtpr: rtpr
- 9: LL: rttr ll
- 10: LM: rttr lm
- 11: ML: rttr ml
- 12: MM: rttr mm

Setting these values has no effect on any other command. However, once they have been set, they remain set until the user changes them with another

"SETOUTPUT" command.

Type

Miscellaneous

SETPARAMETERS

This command prompts the user to set all of TETRAD's parameters. These are, in order:

- Covariance file
- Graph file
- Output file
- TTR of base model
- Edges in base model
- Number of latent variable of 0 indegree in base model
- Units of suggested sets
- Significance level

The following is an example of the use of the command to change the significance level, the units of suggested sets, and the output file.

{ The items in brackets are the default values. Entering a <CR> enters the default value. }

>setpar

Covariance file: [d1.dat] <CR>

Graph file: [d1graph] <CR>

Output file: out2

TTR for the base model: [6.37395] <CR>

How many edges in base model [11.0000]: <CR>

How many latent variables of 0 indegree in base model: [1.00000] <CR>

Units of suggested sets [2.00000]: 3

Significance level (between 0 and 1) [0.05000]: .4

>

Type

Parameter

Affects parameters

Covariances, Current graph, Significance,
 Number of edges in base model, TTR of base model,
 Number of latent variables of indegree 0 in base model

SIGNIFICANCE

The significance command sets the significance level used by TETRAD to determine whether tetrad and partial equations hold in the population. If a well-formed number between 0 and 1 is not entered, then it will continue to prompt until one is entered.

Type

Parameter

Affects parameter

Significance

STANDARD

The Standard command prints out the TTR, I-H, H-I, number of tetrad equations that hold in the population, number of tetrad equations explained by the model, TPR, PI-PH, PH-PI, number of partial equations that hold in the population, and number of partial equations explained by the model.

Type

Output-producing

Parameters that affect

Covariances, Current graph, Significance,
 Number of edges in base model, TTR of base model,
 Number of latent variables of indegree 0 in base model

See also

Input, Changegraph, Significance, Pi

STATUS

Status prints out the current value of each output parameter. The output options refer to parameters set by the "SETOUTPUT" command and used by the "RUN" command. Following is an example of the format.

```
The significance level is: -1.0000
The number of suggested sets units is:      1
The sample size is:      0
Number of edges in base model:      0
TTR of base model:      0.00
TPR of base model:      0.00
```

Output options:

```
Suggested sets:      TRUE  Implied tetrads: FALSE
All tetrads:      TRUE  All partials:      TRUE
Implied partials: FALSE  Held tetrads:      FALSE
Held partials:      FALSE LLRttr:      TRUE
LMRttr:      TRUE  MMRttr:      TRUE
MLRttr:      TRUE  MLRttr:      TRUE
Prttr:      TRUE
```

Type

Miscellaneous

SUGGESTEDSETS

Suggestedsets outputs a series of significance levels. Associated with each significance level is a series of sets of treks. Each set of treks will defeat some tetrad equations that do not hold in the population at the given significance level, but will not defeat any tetrad equation that holds in the population at the given significance level. A heuristic procedure is used to calculate the sets, so it is possible, although unlikely that there are some inaccuracies in the sets output. The number of different significance levels is determined by the units parameter.

The first significance level is the first level at which any suggested sets exist. Each succeeding significance level is the first level at which there are different suggested sets than at the preceding level.

Below is an example of output from the "SUGGESTEDSETS" command with covariance file d1.dat and graph file d1graph.

Sets of suggested treks at significance level = 0.0063

{x4-x5 }

Sets of suggested treks at significance level = 0.0207

{x3-x4 x3-x5 x4-x5 }

If units had been set at one, then only the suggested sets at significance level 0.0063 would have been suggested.

Switches

none

Parameters that affect

Covariances, Current Graph, Units

See also

Input, Changegraph

TETRAD

Tetrad prints out information about a set of tetrad equations. The membership of the set is determined by the switches. For each member of the set tetrad prints out whether it is implied, whether it holds at the given significance level, and the probability of obtaining the observed tetrad residual given that the residual in the population is 0.

Switches

Default switch: All

- All: Information about every quartet of distinct observable variables. If this switch is on, the other switches are ignored.
- Held: Information about every tetrad equation among four distinct observable variables that holds at the given significance level.
- Implied: Information about every implied tetrad equation among four distinct

observable variables.

Type

Output-producing

Parameters that affect

Covariances, Current graph, Significance,
Number of edges in base model, TTR of base model,
Number of latent variables of indegree 0 in base model

See also

Input, Changegraph, Significance, Pi

UNITS

The units command determines how many significance levels the suggestedsets command will print out suggested sets for. If a well-formed number is not entered, it will continue to prompt the user until one is entered. If a real number is entered instead of an integer, it will round the number to the nearest integer.

Type

Parameter

Affects parameter

Units

VIEW

View prompts for a filename. It then prints out the file a screenful at a time. It recognizes the following commands. These commands are executed immediately upon being typed in: the user does not enter a CR.

It is possible to split the screen into up to eight windows. Each window may contain a separate file. It displays only the first 80 characters of a line.

- b Move up one line. (No effect if top of file on screen)
- d Print next screenful. (No effect if bottom of file on screen)
- e Exit view mode, return to TETRAD.
- f Move down one line. (No effect if bottom of file on screen)
- g Get file. (Prompts for file name).
- n Move to next window.
- p Move to previous window.
- r Remove window.
- s Split window.
- u Print last screenful. (No effect if top of file on screen)

Type

Miscellaneous

Appendix I

The Grammar of the Input

The Backus-Naur notation is used to formally specify the syntax of the input to TETRAD. A complete description of this notation can be found in PASCAL. A short summary of the notation appears below.

```

form           : a syntactical object of type form
forma formb    : the concatenation of forma and formb
(forma formb)  : also the concatenation of forma and formb
{form}         : 0 or 1 occurrences of form
form*          : any number (including 0) occurrences of form
form+          : at least one occurrence of form
'a'            : the character a
forma | formb  : an occurrence of either forma or formb
  
```

The correct syntax of the program's input is given below. The following conventions are used.

- The character produced by pressing the space bar is denoted SPACE.
- The character produced by pressing the tab key is denoted TAB.
- The character produced by pressing the return key is denoted CR.

```
space ::= TAB | SPACE
```

```
digit ::= '0' | '1' | '2' | '3' | '4' | '5' | '6' | '7' | '8' | '9'
```

```

alphanumeric ::=
  'a' | 'b' | 'c' | 'd' | 'e' | 'f' | 'g' | 'h' |
  'i' | 'j' | 'k' | 'l' | 'm' | 'n' | 'o' | 'p' |
  'q' | 'r' | 's' | 't' | 'u' | 'v' | 'w' | 'x' |
  'y' | 'z' | 'A' | 'B' | 'C' | 'D' | 'E' | 'F' |
  'G' | 'H' | 'I' | 'J' | 'K' | 'L' | 'M' | 'N' |
  'O' | 'P' | 'Q' | 'R' | 'S' | 'T' | 'U' | 'V' |
  'W' | 'X' | 'Y' | 'Z'
  
```

```
alphanumeric ::= digit | alphanumeric
```

```
sign = '+' | '-'
```

```
number ::= {sign} {digit} '.' digit+
variable ::= alphanumeric {alphanumeric}
edge ::= variable space+ variable
number-line ::= space* number space* CR
edge-line ::= space* edge space* CR
input-line ::= space* number space+ edge-line
variable-line ::= space* variable (space+ variable)* CR
numbersgroup ::= space* number (space+ number)
graph-file ::= edge-line+
line-form ::= input-line+
matrix-form ::= variable-line numbersgroup
covariance-file ::= line-form | matrix-form
```

The covariance file has form covariance-file.

The graph file has form graph-file.

When TETRAD prompts the user for a number, a number line should be entered.

When TETRAD prompts the user for an edge, an edge line should be entered.

ABOUT THE AUTHORS

Clark Glymour took undergraduate degrees in chemistry and philosophy at the University of New Mexico and a doctorate in History and Philosophy of Science from Indiana University, where he also pursued minors in mathematics and chemical physics. He is the author of *Theory and Evidence* (Princeton University Press, 1980), an editor of *Foundations of Space-Time Theories* (University of Minnesota Press, 1981), and a co-author of several historical studies on the development of modern physics. His *Freud's Science* is forthcoming from Harvard University Press. He is presently Professor and Head of the Department of Philosophy at Carnegie-Mellon University, and Adjunct Professor of History and Philosophy of Science at the University of Pittsburgh.

Richard Scheines studied history at Hobart College, and is a Ph.D candidate in History and Philosophy of Science at the University of Pittsburgh. He is Research Programmer for the Center for the Design of Educational Computing at Carnegie-Mellon University.

Peter Spirtes took degrees in physics and in philosophy at the University of Michigan, and both a doctorate in History and Philosophy of Science and a master's degree in Computer Science from the University of Pittsburgh. He is now Research Programmer in the Laboratory for Computational Linguistics, and Research Scientist in the Department of Philosophy at Carnegie-Mellon.

Kevin Kelly took an undergraduate degree in philosophy at the University of Missouri-St. Louis and then received a doctorate in History and Philosophy of Science from the University of Pittsburgh. He is the author of several papers on computation theoretic aspects of computer discovery procedures. He is Assistant Professor of Philosophy at Carnegie-Mellon University.

REFERENCES

- [1] Aho, A., Hopcroft, J., Ullman, J.
The Design and Analysis of Computer Algorithms.
Addison-Wesley Publishing Company, 1974.
- [2] Anderson, T.
An Introduction to Multivariate Statistical Analysis.
Wiley, New York, 1984.
- [3] Baumrind, D.
Specious Causal Attributions in the Social Sciences.
Journal of Personality and Social Psychology 45:1289-1298, 1983.
- [4] Bentler, P. and Woodward, J.
A Head Start Reevaluation: Positive Effects Are Not Yet Demonstrable.
Evaluation Quarterly 2:493-510, 1978.
- [5] Bentler, P., Bonett, D.
Significance Tests and Goodness of Fit in the Analysis of Covariance Structures.
Psychological Bulletin 88:588-606, 1980.
- [6] Bentler, P.
Theory and Implementation of EQS: A Structural Equations Program
BMDP Statistical Software Inc., 1964 Westwood Boulevard, Suite 202, LA, CA
90025, 1985.
- [7] Blalock, Hubert M.
Causal Inferences in Nonexperimental Research.
The University of North Carolina Press, Chapel Hill, 1961.
- [8] Blalock, Hubert M.
Theory Construction: From Verbal to Mathematical Formulations.
Prentice-Hall, Englewood Cliffs, N.J., 1969.
- [9] Blalock, Hubert M.
Conceptualization and Measurement in the Social Sciences.
Sage Publications, Beverly Hills, 1982.
- [10] Blau, Peter; Duncan, Otis Dudley.
The American Occupational Structure.
Wiley, New York, 1967.
- [11] Blum, R.
Discovery, Confirmation and Incorporation of Causal Relationships from a Large
Time-oriented Clinical Data Base: The RX Project
In W. Clancy and E. Shortliffe (editors), *Readings in Medical Artificial
Intelligence*, pages 399-425. Addison-Wesley, Reading, Mass., 1984.
- [12] Brophy, J. and Good, T.
Teacher-Student Relationships: Causes and Consequences.
Holt, Rinehart and Winston, New York, 1974.
- [13] Byron, R.
Testing for Misspecification in Econometric Systems Using Full Information.
International Economic Review 28:138-151, 1972.
- [14] Campbell, D., Schwartz, R, Sechrest, L., and Webb, E.
Unobtrusive Measures: Nonreactive Research in the Social Sciences.
Rand McNally, Chicago, 1966.

- [15] Cicirelli, V. et al.
Report to the Office of Economic Opportunity. Volume I & II: The Impact of Head Start: An Evaluation of the Effects of Head Start on Children's Cognitive and Affective Development.
Ohio University and Westinghouse Learning Corporation, Athens, Ohio, 1969.
- [16] Chudde, C., McCrone, D.
The Linkage Between Constituency Attitudes and Congressional Voting Behavior: A Causal Model.
American Political Science Review 60:66-72, 1966.
- [17] Costner, H., Schoenberg, R.
Diagnosing Indicator Ills in Multiple Indicator Models.
In Goldberger, A., Duncan, O. (editors), *Structural Equation Models in the Social Sciences*. Seminar Press, New York, 1973.
- [18] Costner, H. and Herting, J.
Respecification in Multiple Indicator Models.
1984.
- [19] Crain, W.
Vehicle Safety Inspection Systems: How Effective?
American Enterprise Institute for Public Policy Research, Washington, 1979.
- [20] Cross, C.
Studies in Modality.
PhD thesis, University of Pittsburgh, 1985.
Department of Philosophy.
- [21] Efron, B.
Bootstrap Methods: Another Look at the Jackknife.
Annals of Statistics 7:1-26, 1979.
- [22] Ehrenberg, A. S. C.
Data Reduction.
Wiley, New York, 1975.
- [23] Ericksen, E and Kadane, J.
Estimating the Population in a Census Year.
Journal of the American Statistical Association 80:98-131, 1985.
- [24] Faust, David.
The Limits of Scientific Reasoning.
University of Minnesota, Minneapolis, 1984.
- [25] Forbes, H., Tufte, E.
A Note of Caution In Causal Modelling.
American Political Science Review 62:1258-1264, 1968.
- [26] Fox, J.
Linear Statistical Models and Related Methods.
John Wiley and Sons, New York, 1984.
- [27] Freedman, D. and Navidi, W.
Regression Models for Adjusting the 1980 Census.
Statistical Science, forthcoming.

- [28] Freedman, D.
Structural-Equation Models: A Case Study.
Technical Report, University of California, Berkeley, May, 1983.
Department of Statistics.
- [29] Freund, John. E.
Mathematical Statistics.
Prentice-Hall, Inc., 1971.
- [30] Gale, W. and Pregibon, D.
Artificial Intelligence Research in Statistics.
The AI Magazine -:72-75, Winter, 1985.
- [31] Garnett, J. C. M.
On Certain Independent Factors in Metal Measurements.
Proceedings of the Royal Society 96:xx, 1919.
- [32] Giere, R.
Bayesian Statistics and Biased Procedures.
Synthese 20:371-387, 1969.
- [33] Glymour, C.
The Epistemology of Geometry.
Nous , 1977.
- [34] Glymour, C.
Theory and Evidence.
Princeton University Press, 1980.
- [35] Glymour, C.
Social Science and Social Physics.
Behavioral Science 28:xx, 1983.
- [36] Glymour, C.
Statistics and Metaphysics.
1985.
Lecture to the American Statistical Association.
- [37] Good, I. J.
A Causal Calculus, I, II.
British Journal For Philosophy of Science 11 and 12:305-318 and 43-51, 1961
and 1962.
- [38] Gorry, G. and Barnett, G.
Experience with a Model of Sequential Diagnosis.
Computers and Biomedical Research 1:490-507, 1968, 1968.
- [39] Granger, C.
Investigating Causal Relations By Econometric Models and Cross-Spectral Methods.
Econometrica 37:424-438, 1969.
- [40] Guilford, J.
Psychometric Methods.
McGraw Hill, New York, 1936.
- [41] Harary, F. and Palmer, E.
Graphical Enumeration.
Academic Press, New York, 1973.

- [42] Hausman, D.
Are There Causal Relations Among Dependent Variables?
Philosophy of Science 50:58-81, 1983.
- [43] Heise, D.
Causal Analysis.
John Wiley & Sons, New York, 1975.
- [44] Hempel, Carl G.
Aspects of Scientific Explanation.
The Free Press, New York, 1965.
- [45] Holland, Paul,
Statistics and Causal Inference.
1985.
Lecture to the American Statistical Association, 1985.
- [46] Holzinger K. and Spearman, C.
Note on the Sampling Error of Tetrad Differences.
British Journal of Educational Psychology 20:91-97, 1926.
- [47] Holzinger, K.
Statistical Resume of the Spearman Two-factor Theory.
University of Chicago Press, Chicago, 1930.
- [48] James, L. et al.
Causal Analysis: Assumptions, Models and Data.
Sage Publications, Beverley Hills, 1982.
- [49] Jeffreys, H.
Scientific Inference.
Cambridge University Press, Cambridge, 1957.
- [50] Joreskog, K. and Sorbom, D.
LISREL IV - User's Guide.
International Educational Services, Chicago, 1978.
- [51] Magidson, J. (editor).
Advances in Factor Analysis and Structural Equation Models.
Abt Books, Cambridge, 1979.
- [52] Joreskog, K. and Sorbom, D.
LISREL VI User's Guide
third edition, Scientific Software, Inc., Mooresville, Indiana, 1984.
- [53] Kadane, J.
Reply to Freedman and Navidi.
Statistical Science, forthcoming.
- [54] Kahneman, D., Slovic, P., Tversky, A. (editors).
Judgement Under Uncertainty: Heuristics and Biases.
Cambridge University Press, 1982.
- [55] Kelley, T.
Crossroads in the Mind of Man.
Stanford University Press, Stanford, 1928.
- [56] Kenny, D.
Correlation and Causality.
Wiley, New York, 1979.

- [57] Kmenta, J., Ramsey, J. (editors).
Large-Scale Macro-Economic Models.
North-Holland, Amsterdam, 1981.
- [58] Kohn, M.
Class and Conformity.
Dorsey Press, Homewood, Il., 1969.
- [59] Leamer, E.
Let's Take the Con Out of Econometrics.
American Economic Review 73:31-43, 1983.
- [60] Lee, S.
Analysis of Covariance and Correlation Structures.
Computational Statistics & Data Analysis 2:279-295, 1985.
- [61] Levi, Isaac.
*Borzoï Books in the Philosophy of Science: Gamblin With Truth; An Essay on
Induction and the Aims of Science*.
Knopf, New York, 1967.
- [62] Lewis, D.
Counterfactuals.
Harvard University Press, Cambridge, 1973.
- [63] Lewis, D.
Causation.
Journal of Philosophy 70, 1973.
- [64] Lieberman, S.
Making It Count.
University of California Press, Berkely, CA, 1986.
- [65] Lindley, D., Tversky, A. and Brown, R.
On the Reconciliation of Probability Assessments.
Journal of the Royal Statistical Society 142:146-180, 1979.
- [66] Lindsay, R., Buchanan, B., Feigenbaum, E. and Lederberg, J.
Applications of Artificial Intelligence For Organic Chemistry.
McGraw Hill, New York, 1980.
- [67] Ling, Robert.
Review of "Correlation and Causation" by David Kenny.
Journal of the American Statistical Association 77:489-491, 1983.
- [68] Mackie, J.
The Cement of the Universe.
Oxford University Press, Oxford, 1974.
- [69] Magidson, J.
Toward a Causal Model Approach for Adjusting for Preexisting Differences in the
Nonequivalent Control Group Situation.
Evaluation Quarterly 1:399-420, 1977.
- [70] Magidson, J. and Sorbom, D.
Adjusting for Confounding Factors in Quasi-Experiments: Another Reanalysis of the
Westinghouse Head Start Evaluation.
Educational Evaluation and Policy Analysis 4:321-329, 1982.

- [71] Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J. Parzen, E., and Winkler, R.
The Forecasting Accuracy of Major Time Series Methods.
John Wiley & Sons, New York, 1984.
- [72] Maruyama, G. and McGarvey, B.
Evaluating Causal Models: An Application of Maximum Likelihood Analysis of Structural Equations.
Psychological Bulletin 87:502-512, 1980.
- [73] McC. Adams, et al. (editor).
Behavioral and Social Science Research: A National Resource.
National Academy Press, Washington, D.C., 1982.
- [74] McManus.
Bayesian Estimation of the Deterrent Effect of Capital Punishment.
mimeo, University of California, Los Angeles, 1981.
- [75] Meehl, P.
Clinical versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence.
University of Minnesota Press, Minneapolis, 1954.
- [76] Meehl, P.
Theoretical Risks and tabular Asterisks: Sir Karl, Sir Ronald, and the Slow Progress of Soft Psychology.
Journal of Counseling and Clinical Psychology 46:806-834, 1978.
- [77] Miller, Warren, and Stokes, Donald.
Constituency Influence in Congress.
American Political Science Review 1963:45-56, March, 1963.
- [78] Miller, J., Slomczynski, K. and Schoenberg, R.
Assessing Comparability of Measurement in Cross-National Research: Authoritarian-Conservatism in Different Sociocultural Setting.
Social Psychology Quarterly 44:178-191, 1981.
- [79] O'Donnell, J. and Clayton, R.
The Stepping-stone Hypothesis--Marijuana, Heroin and Causality.
Chemical Dependencies: Behavioral and Biomedical Issues 4:229-241, 1982.
- [80] Pearson, K. and Moul, M.
The Mathematics of Intelligence, I. The Sampling Errors in the Theory of a Generalized Factor.
Biometrika 19:246-291, 1927.
- [81] Rosenthal, R.
Experimental Effects in Behavioral Research.
Appleton-Century Crofts, New York, 1966.
- [82] Rossi, Peter; Berk, Richard; Lenihan, Kenneth.
Money, Work and Crime.
Academic Press, New York, 1980.
- [83] Rossi, P., Berk, R., and Lenihan, K.
Saying It Wrong with Figures: A Comment on Zeisel.
American Journal of Sociology 88:390-393, 1982.

- [84] Salmon, Wesley C.
Probabilistic Causality.
Pacific Philosophical Quarterly 61:50-74, 1980.
- [85] Savage, L.
The Foundations of Statistics.
Wiley, New York, 1954.
- [86] Scheffe, H.
The Analysis of Variance.
Wiley, New York, 1959.
- [87] Schoenberg, R. and Richtand, C.
Application of the EM Method.
Sociological Methods and Research 13:127-150, 1984.
- [88] Schultz, T.
Monographs of the Society for Research in Child Development. Volume 47:
Rules of Causal Attribution.
x, x, 1982.
- [89] Searle, S. R.
Linear Models.
Wiley, New York, 1971.
- [90] Simon, H.
Causal Ordering and Identifiability.
In *Models of Discovery*, pages 53-80. D. Reidel, Dordrecht, 1953.
- [91] Simon, H.
Spurious Correlation: A Causal Interpretation.
In *Models of Discovery*, pages 93-106. D. Reidel, Dordrecht, 1954.
- [92] Simon, Herbert A.
Models of Discovery.
D. Reidel Publishing Company, 1977.
- [93] Langley, P., Simon, H., Bradshaw, G., Zytkow, J.
Scientific Discovery: An Account of the Creative p Process.
MIT Press, Boston, 1986.
- [94] Sims, C.
Money, Income, and Causality.
American Economic Review 62:540-552, 1972.
- [95] Skinner, B. F.
About Behaviorism.
Vintage Books, New York, 1976.
- [96] Skyrms, Brian.
Causal Necessity.
Yale University, New Haven, 1980.
- [97] Sorbom, D.
Structural Equation Models with Structured Means.
In Joreskog, K. and Wold, H. (editors), *Systems Under Indirect Observation:
Causality, Structure, and Prediction*. North-Holland, Amsterdam, 1981.
- [98] Sorbom, D.
Detection of Correlated Errors in Longitudinal Data.
British Journal of Mathematical and Statistical Psychology 28:138-151, 1975.

- [99] Spearman, C.
The Abilities of Man.
Macmillan, New York, 1927.
- [100] Spearman, C.
Pearson's Contribution to the Theory of Two Factors.
British Journal of Psychology 19:95-101, 1928.
- [101] Spearman, Charles.
General Intelligence Objectively Determined and Measured.
American Journal of Psychology 15:201-293, 1904.
- [102] Glymour, C. and Spirtes, P.
Space-time and Synonymy.
Philosophy of Science 39:xx, 1983.
- [103] Suppes, Patrick.
A Probabilistic Theory of Causality.
North-Holland, Amsterdam, 1970.
- [104] Thurstone, L.
The Vectors of Mind.
University of Chicago Press, Chicago, 1935.
- [105] Timberlake, Michael, and Williams, Kirk R.
Dependence, Political Exclusion, and Government Repression: Some Cross-National Evidence.
American Sociological Review 49:141-146, 1984.
- [106] Van Fraassen, Bas C.
The Scientific Image.
Clarendon Press, 1980.
- [107] Wheaton, B., Muthen, B., Alwin, D. and Summers, G.
Assessing Reliability and Stability in Panel Models.
In D. Heise (editor), *Sociological Methodology 1977*, pages 84-136. Jossey-Bass, San Francisco, 1977.
- [108] Wishart, J.
Sampling Errors in the Theory of Two Factors.
British Journal of Psychology 19:180-187, 1928-29.
- [109] Wright, S.
The Method of Path Coefficients.
Annals of Mathematical Statistics 5:161-215, 1934.
- [110] Zeisel, Hans.
Disagreement Over the Evaluation of a Controlled Experiment.
American Journal of Sociology 88:378-389, 1982.
- [111] Zellner, A.
Philosophy and Objectives of Econometrics.
In Currie, D., Nobay, R., and Peel, D. (editors), *Macroeconomic Analysis: Essays in Macroeconomics and Econometrics.* Croom Helm Publishing Co., London, 1981.

Index

- (H-I) 105, 115
- (I-H) 105, 115
- * (special operation) 306
- 8087 co-processor 338
- AC personality trait 140
- Acceptance of hypotheses 28
- Alien.dat 239
- Alienation 239
- Anchor.dat 173
- Anchoring effect 171
- Anderson, T. 108
- Anomia 239
- Anti-realism 28
- Approximate truth 128
- Approximation in science 127
- Artificial intelligence 15, 17
 - aims of 17
- Atomic theory 65
- Authoritarianism-conservatism case
 - variable abbreviation scheme 153
- Avogadro's number 41
- Background knowledge 253
- Baumrind, D. 26, 31, 35, 36
- Bayesian objections (to mechanized discovery) 49
- Bayesians 261
- Beam search 129, 151
- Belief in hypotheses 28
- Bentler, P. 110, 111, 210, 239, 240
- Berks, R. 32
- Binet 261
- Blalock, H.M. 16, 30, 81, 197, 271
- Blau, P. 27, 37, 54
- Blum 18
- Bohr, N. 29
- Bonett, D. 110, 111
- Boole, G. 261
- Boyle, R. 34
- Breadth first search 104
- Brophy, J. 227
- Buchanan, B. 19
- Byron, R. 16, 128
- Campbell, D. 171, 271
- Cannizzarro, S. 41, 65
- Capital punishment 14
- Case studies
 - alienation 131, 239
 - authoritarianism-conservatism 140, 170
 - congressional voting behavior 197
 - Head Start 210
 - SAT Scores 193
 - school achievement 224
 - simulated 254
 - SYSTAT BASIC 134
- Causal connections
 - initial bias against 101
 - kinds of 11
- Causal inference 24
- Causal inferences
 - weak version 24
- Causal modelling
 - a Bayesian approach to 15
- Causal relations
 - graphic representation of 24
- Causal relationship
 - vs. correlated errors 143
- Causal theory 11

- Causality
 - as mechanism 27
 - as oomph 27
 - counterfactual account 27
 - probabilistic account 27
- Causation and functional dependence 24
- Celestial mechanics 36
- Census undercount 13, 23
- Certainty 29
- Chain (of equations) 313
- Chi square test 97
- CI (Cutright index) 152
- Cicirelli, V. 210
- Circular chain (of equations) 313
- Cnudde, C. 197
- Compare function 105, 115
- Confirmation 65
- Connected variables 143
- Connection
 - of source to sink 288
- Containment (of vertices in paths) 71
- Coordinating path effect 80, 110
- Copernicus, N. 36, 65, 264
- Correlated error terms 269
- Correlation 79
- Correlation coefficient 263
- Correlation residual 106
- Costner and Schoenberg's procedure 129
- Costner, H. 128, 129, 152, 156, 272
- Covariance 79
 - fully expanded form of 287
- Covariance file 154
- Covariance matrix
 - overidentifying constraints on 267
- Crain, W. 27
- Credence 29
- Cross, C. 23
- Cutright index 152
- Cycle 12, 288

- D(I-H) 103, 119
- Dalton, J. 29, 45
- Darwin, C. 36
- Data analysis
 - confirmatory 53
 - exploratory 53
- Degenerate edge 288
- Degenerate path
 - length of 288
- Degrees of freedom
 - vs. unexplained residuals 112
- DENDRAL 64
- Dependent variable
 - first level expansion of 291
- Digraph 288
 - acyclic 288
 - labelled 289
- Direct descendence (among variables) 293
- Directed edge (in a graph) 71
- Directed graph 70
- Divide and conquer strategy 279
- DOS 338
- DP ("dontpick" set) 279, 284
- Duhem, P. 28
- Dulong, P. 44
- Dumas, J. 34
- Duncan, O. 27, 37, 54, 272

- Eddington, A. 45, 264
- Edge 288
 - from 288
 - head of 288
 - into 288
 - out of 288
 - Rtrr value of 284

- tail of 288
- Ehrenberg, A.S.C. 36
- Einstein, A. 64
- Energy 152
- Entailment of constraints 83
- Eotvos experiment 64
- EQS 1, 78, 97, 127, 131, 151, 239, 246, 327, 338
 - in batch job 338
 - running batch jobs 339
- Equation coefficient 289
- Equationals
 - consistent, non-redundant set of 289
- Equivalent subgraphs 280
- Ericksen, E. 13
- Evidence
 - tainted 65
- Exec 152
- Expansions
 - covariances 293
 - products of covariances 293
 - variances 293
- Experiment 11
- Experimentation 7
 - ambiguity of 35
 - combinatoric significance of 31
- Explanation 10, 16, 30, 65, 79, 248
 - and reduction of residuals 111
 - of constraints 202
 - of correlation in the data 72
 - of tetrad equations 262
 - of variance (probability of) 111
 - robust 264, 271
 - without latent variables 265
- Explanatory power 50, 128

- Factor analysis 16, 100
 - introduction of 269
- Faust, D. 10
- First level expansion of a dependent variable 291
- Fisher's z statistic 109
- Fisher, R. 51
- FIXEDX 334
- Forbes, H. 197
- Fox, J. 97
- Freedman, D. 36, 38
- Full expansion of a variable 291

- Gale 18
- Galton, F. 261, 270
- Garnett, M. 267
- Gaudin, A. 34
- Gay-Lussac 45
- Glymour, C. 22, 41
- GM 284
- GNP 152
- Good, I.J. 24
- Good, T. 227
- Granger, C. 24
- Graph
 - acyclic 72
 - cyclic 72
 - directed 10, 69
 - disconnected 90
 - statistical properties of 69
 - subgraph 90
 - Tetrad-compatible 305
 - vertices of 10
 - without cycles 12
- Graph file 155
- Guilford 270

- Hausman, D. 23
- Head (of an equation) 313
- Head Start 15, 23, 30, 151
- Heise, D. 83, 273

- Herting, J. 129
 Hertz, P. 45
 Heuristic search 17
 Holland, P. 22, 25, 29
 Holzinger, K. 265, 267, 269, 270, 272
 Hotelling, H. 269
- I(H-I) 103, 119, 145
 Ideal gas law 44
 Identical equality of terms 293
 Implication
 of constraints 83
 robust 269, 271
 Ind 306
- James, L. 111
 Jeffreys, H. 16, 50, 264
 Joreskog, K. 131, 151, 193, 239, 240
- Kadane, J. 13
 Kahneman, D. 171
 Kelley, T. 265, 267, 270, 272
 Kenny, D. 36, 54
 Kepler, J. 29, 36, 264
 Keynes, J.M. 15
 Kmenta, J. 10
 Kohn, M. 115, 140, 170
 Kuhn, T. 9
- Label of zero length path 276
 Labor 152
 Latent variables
 when to introduce 193
 Leamer, E. 15, 31
 Least squares loss function 111
 Lewis, D. 23, 27
 Lieberman, S. 25
 LIFE 32
 Likelihoods 127
 probabilities of test statistics as 127
 Ling, R. 36, 54
 LISREL 1, 16, 78, 97, 112, 127, 131, 151, 212, 226, 239, 272, 327
 inadequacy of 129
 modelling restrictions circumvented 330
 restrictions on causal connection 328
 variable types 328
 LISREL variables
 abbreviation scheme 328
 LISREL VI 129
 modification indices 129
 LM_i 284
 Logical omniscience 49
 Loop 305
 Loop form 306
 Loop trek form 306
 Loops touching open paths 305
- Mackie, J.L. 27
 Magidson, J. 15, 210
 Makridakis, S. 10
 Maruyama, G. 31, 224
 Mason's rule 305
 Mason, S. 273
 McCrone, D. 197
 McGarvey, B. 31, 224
 McManus 14
 Medical diagnosis 49
 Meehl, P. 9, 21, 109
 Methodological virtues
 as restrictions on prior distributions 50
 as utilities 50
 Michelson-Morley experiment 34
 Miller, D.C. 34
 Miller, J. 170
 Miller, W. 197, 198

- Model 10
 - best elaboration of 99
- Models
 - as approximations 38
 - causal 69, 88
 - circuits in 273
 - degrees of freedom of 74
 - econometric 22, 69
 - factor analytic 18, 69
 - heteroscedastic 79
 - nested sequences of 16
 - non-recursive 273
 - path analytic 18, 69
 - regression 69, 97
 - statistical 70, 87
 - structural equation 18, 69
 - structural equation (defined) 79
- Modification indices 129
- Moments 88
- Monte Carlo Data 135
- Monte Carlo trials 61
- Moul, M. 266

- N level expansion of a dependent variable 291
- N+1 level expansion of a dependent variable 291
- National Survey 115, 151, 170
- Newton, Sir Isaac 29, 45
- Non-redundancy (of equation sets) 289
- Normal forms for covariances
 - loop 306
 - loop trek 306
 - TETRAD trek 307
- Normed fit index 110
- Normed residuals 128

- O'Donnell, J. 26
- Octad differences 90
- Open path 305
- Overconfidence 64
- Overidentifying constraints 70, 81, 267

- Parameter estimation 97, 271
 - least squares 97
 - maximum likelihood 97
- Parameters
 - unspecified 70
- Parsimony 111
- Partial correlation residual 107
- Partial correlations 82, 102, 275
 - vanishing 82
- Path 71
 - acyclic 72
 - as causal influence 72
 - correlation due to 72
 - covariance due to 72
 - cyclic 72
 - degenerate 288
 - label of 276
 - loop 305
 - open 305
 - sink of 288
 - source of 288
- Path finding algorithm 276
- Path of length n 288
- Path search algorithm
 - complexity of 277
- Paths
 - containing cycles 288
 - intersection of 288
 - open 276
- Pearson, K. 261, 270
- Pentad differences 91
- Pentad equations 165
- Perrin, J. 41
- Petit, A. 44

- Phi (LISREL variable type) 328
- Pi value 105, 115, 158
- Point of intersection (of paths) 288
- Population correlation matrix 82
- Post-Enumeration Program 13
- Pound 36
- Powerlessness 239
- PR value 107
- Prediction 65
- Prediction equations 81
- Pregibon 18
- Prior knowledge of the user 151
- Psi (LISREL variable type) 328
- Ptolemy 65, 264, 270

- RAM requirement 1
- Ramsey, F. 10
- Random variables 289
- Realism 28
- Rebka 36
- Relevance of loops to variables 306
- Reliability 59
- REX 18
- Richtand, C. 170
- RL_{ij} 306
- Rosenthal, R. 35
- Rossi, P. 14, 32
- Rtpr chart 105, 115
- Rttr chart 105, 115, 119, 145, 157
 - time to compute 284
- Rttr value (of a trek) 284
- RX 18

- Sample correlation matrix 82
- Sample space 289
- Sampling distribution
 - of products of tetrad residuals 265
- Sampling error 265
- Sat.dat 193
- Satisficing 17
- Savage, L.J. 50
- Scheffe', H. 109
- Schoenberg, R. 128, 152, 156, 170, 272
- Science
 - as simplification 269
- Scientific discovery 17
- Search strategies for improved models 103
 - breadth-first search 104
- Sellars, W. 17
- Semi-ring
 - closed 276
- SES 239
- Sextad differences 90
- Shultz, T. 26
- Significance level 158
- Simon, H. 17, 19, 37, 75, 109, 197, 271
- Simon-Bialock technique 200
- Simplicity 50
- Simplicity Principle 101
- Sims, C. 24
- Skinner, B.F. 29
- Skyrms, B. 24
- SLCT 289
 - acceptable 290
 - narrow sense of 291
- SLCT's
 - isomorphism between 291
- Socio-economic status 239
- Sorbom, D. 16, 128, 131, 151, 210, 239, 240
- Spearman's Principle 100, 101, 121, 130, 206, 245, 284
- Spearman, C. 261, 265, 270
- Statistical diagnostics 18
- Statistical estimator 56
- Statistical inference 17
- Statistical model 10

- Statistical modelling
 - underlying assumptions of 11
- Statistical tests 56
 - Neyman-Pearson theory of 127
- Stochastic linear causal theory 289
- Stokes, D. 197, 198
- Strong implication 287
- Substantive knowledge 12
- Suggested Sets (of trek additions) 114, 156, 277
- Sum of squares of residuals 111
- Suppes, P. 24
- SYSTAT BASIC 134, 254

- T₀ 307
- T_{ij}
- Tail (of an equation) 313
- TARP 32, 37
- Term coefficient 287
- Test
 - chi square 53
 - Neyman-Pearson account of 59
- Testing null hypotheses 109
- TETRAD 16, 18, 19, 28, 30, 46, 50, 53, 62, 69, 78, 85, 91, 97, 104, 108, 110, 135, 141, 143, 144, 261, 273
 - algorithm of 104
 - assumptions for use 114
 - automatic search procedure of 112
 - inputs to 104
 - output information 105
- Tetrad difference 101
 - sampling distribution of 108
 - sampling variance of 108
 - variance of 266
 - when correlations are small 109
- Tetrad equations 82, 85, 105, 114, 263, 275
 - calculation of 110
 - calculation rule 276
 - intractable number of 271
 - number of 108
 - statistical dependence of 109
- Tetrad residual 265
 - normed fit index for 111
 - sampling distribution of 265
- Tetrad residuals
 - sum of (c.f. TTR value) 112
- TETRAD trek coefficient 307
- TETRAD trek form 307
- Tetrad-compatibility 305
- Theory 10
- Theta delta (LISREL variable type) 328
- Theta epsilon (LISREL variable type) 328
- Thomson, G. 264
- Thorndike 264
- Thurstone's Principle 100, 101, 103, 107, 120, 206, 245
- Thurstone, L. 269
- Timberlake 188
- TPR value 105, 107, 115
- Transitional Aid Research Project 32
- Translating graphs into equations 76
- Treatment effects 35
- Trek 72, 99, 113, 288
 - open 72
 - product 314
 - source of 288
 - sum 276
 - ways to introduce 72
- Treks
 - globally minimal sets of 284
 - locally minimal sets of 284
 - open 110, 276, 306
 - three ways to produce 277
- TTR value 105, 107, 115, 145, 156, 284
- Tufte, E. 197
- Tversky, A. 171

- Unexplained residuals
 - vs. increased degrees of freedom 112

- Vanishing partial correlations 105, 114
 - sampling distribution of 109
- Variable
 - full expansion of 291
- Variable expansion
 - collected 293
 - distributed 293
- Variables
 - autocorrelation among 42
 - dependent 22, 69
 - endogenous 22, 69
 - error 73
 - exogenous 22, 287, 69
 - independence property for 289
 - independent 22, 69
 - latent 28, 91, 134
 - maximum number the included program can process 1
 - partial order on 22
 - temporal order of 12
 - time lagged 271
 - unobservable 91
- Vertex
 - indegree of 288
 - outdegree of 288
- Vertices 288
 - adjacent 71
 - indegree of 71
 - outdegree 71
- Wheaton, B. 129, 132, 151, 239
- Williams 188
- Wishart's formula 108
- Wishart, J. 51, 266
- Wold, H. 271
- Woodward, J. 210
- Wright, S. 265, 272
- Yule, G.U. 261
- Zeisel, H. 14, 32
- Zellner, A. 50